SPARC Data Initiative: Comparison of water vapor climatologies from international satellite limb sounders

M. I. Hegglin[1], S. Tegtmeier[2], J. Anderson[3], L. Froidevaux[4], R. Fuller[4], B. Funke[5], A. Jones[6], G. Lingenfelser[7], J. Lumpe[8], D. Pendlebury[6], E. Remsberg[7], A. Rozanov[9], M. Toohey[2], J. Urban[10], T. von Clarmann[11], K. A. Walker[6], R. Wang[12], and K. Weigel[9]

[1]Department of Meteorology, University of Reading, Reading, United Kingdom

[2]GEOMAR, Helmholtz Centre for Ocean Research Kiel, Kiel, Germany

[3]Atmospheric Science, Hampton University, Hampton, VA, United States

[4]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, United States

[5]Instituto de Astrofísica de Andalucía, CSIC, Granada, Spain

[6]Department of Physics, University of Toronto, Toronto, Canada

[7]NASA Langley Research Center, Hampton, VA, United States

[8]Computational Physics, Inc., Boulder, CO, United States

[9]Institute of Environmental Physics (IUP), University of Bremen, Bremen, Germany

[10]Department of Earth and Space Sciences, Chalmers University of Technology, Gôteborg, Sweden

[11]Karlsruhe Institute of Technology, IMK, Karlsruhe, Germany

[12]School of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, GA, United States

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/jgrd.50752

Abstract. Within the SPARC Data Initiative, the first comprehensive assessment of the quality of 13 water vapor products from 11 limb-viewing satellite instruments (LIMS, SAGE II, UARS-MLS, HALOE, POAM III, SMR, SAGE III, MIPAS, SCIAMACHY, ACE-FTS, and Aura-MLS) obtained within the time period 1978-2010 has been performed. Each instrument's water vapor profile measurements were compiled into monthly zonal mean time series on a common latitude-pressure grid. These time series serve as basis for the 'climatological' validation approach used within the project. The evaluations include comparisons of monthly or annual zonal mean cross-sections and seasonal cycles in the tropical and extra-tropical upper troposphere and lower stratosphere averaged over one or more years, comparisons of interannual variability, and a study of the time evolution of physical features in water vapor such as the tropical tape recorder and polar vortex dehydration. Our knowledge of the atmospheric mean state in water vapor is best in the lower and middle stratosphere of the tropics and mid-latitudes, with a relative uncertainty of ±2-6% (as quantified by the standard deviation of the instruments' multi-annual means). The uncertainty increases towards the polar regions (±10-15%), the mesosphere (±15%), and the upper troposphere/lower stratosphere below 100 hPa (±30-50%), where sampling issues add uncertainty due to large gradients and high natural variability in water vapor. The minimum found in multi-annual (1998-2008) mean water vapor in the trop-

ical lower stratosphere is 3.5 ppmv (±14%), with slightly larger uncertainties for monthly mean values. The frequently used HALOE water vapor dataset shows consistently lower values than most other datasets throughout the atmosphere, with increasing deviations from the multi-instrument mean below 100 hPa in both the tropics and extra-tropics. The knowledge gained from these comparisons and regarding the quality of the individual datasets in different regions of the atmosphere will help to improve model-measurement comparisons (e.g. for diagnostics such as the tropical tape recorder or seasonal cycles), data merging activities, and studies of climate variability.

1. Introduction

Water vapor is the most important natural greenhouse gas in the atmosphere and is a positive feedback of the CO_2 climate forcing [*IPCC*, 2007]. The greenhouse effect of water vapor (given a change in absolute values) is strongest in the upper troposphere and lower stratosphere (UTLS) [*Forster and Shine*, 2002], where the lowest temperatures are found and strong gradients across the tropopause region exist [e.g., *Gettelman et al.*, 2011]. Water vapor is also a key constituent in atmospheric chemistry. It is the source of hydroxyl (OH), which controls the lifetime of shorter-lived pollutants, tropospheric and stratospheric ozone, and other longer-lived greenhouse gases such as methane [*Seinfeld and Pandis*, 2006]. Furthermore, water vapor has an important influence on stratospheric chemistry through its ability to form ice, which offers a surface for heterogeneous chemical reactions involved in the destruction of stratospheric ozone [*Solomon*, 1999]. Accurate knowledge of the water vapor distribution and its trends from the upper troposphere up to the mesosphere is therefore crucial to understand climate and chemical forcings.

It is noteworthy that despite the importance of water vapor, there seems to be only little skill in representing water vapor distributions in current chemistry-climate models (used for the assessment of ozone depletion and chemistry-climate interactions) in both the tropical [*Gettelmann et al.*, 2010] and the extra-tropical UTLS [*Hegglin et al.*, 2010], as well as in climate models such as used for the IPCC climate assessments [*Jiang et al.*, 2012] and reanalyses [*Jiang et al.*, 2010] in these regions. The lack of progress in representing UTLS water vapor in models may partially be explained by inconclusive observational records, to which the models are compared [*SPARC CCMVal*, 2010].

Hitherto, observed long-term changes in stratospheric water vapor remain somewhat of a puzzle [*Randel et al.*, 2004; *Hurst et al.*, 2011; *Fueglistaler*, 2012], since the processes that are expected to control water vapor in the stratosphere such as methane oxidation and tropical tropopause temperatures do not show a long-term behavior consistent with the derived water vapor trends. Progress in our understanding is compromised by the current limited knowledge of data quality for both temperature [*Seidel et al.*, 2011] and water vapor measurements. It is not trivial to accurately measure water vapor, and satellite measurements, as well as in situ correlative data, have been shown to exhibit large absolute differences [*SPARC WAVAS*, 2000]. In particular, the current lack of an accepted standard from in situ correlative data precludes a conclusive assessment of the performance of available satellite water vapor measurements [see *Weinstock et al.*, 2009]. Also, while a large number of studies exist on the validation of individual water vapor products [e.g., *Lambert et al.*, 2007; *Milz et al.*, 2009], these focus on profile-by-profile comparisons and have been limited to a subset of the available satellite data.

Within the SPARC Data Initiative [*SPARC Data Initiative*, in preparation; *Hegglin et al.*, in preparation (a)] we have performed the first comprehensive comparison of water vapor products from most of the available limb-viewing satellite instruments from the international space agencies CSA, ESA, NASA, and the Swedish National Space Board (Section 2). Monthly zonal mean water vapor time series within the time frame 1978-2010 have been compiled over the lifetime of each instrument. These climatologies are then compared using a 'climatological' validation approach, rather than the classical validation method based on measurement coincidences (see Section 3). Using climatologies instead of the often small numbers of available coincident profiles may substantially reduce random

errors and hence improve validation [e.g. *Hegglin et al.*, 2008]. Evaluations include the zonal mean atmospheric distribution of water vapor and its seasonality, and other physical features such as inter-annual variability, the tropical tape recorder, or polar vortex dehydration (Section 4). While the 'best' performing instrument is not readily determined, the work presented here is intended to give an overview of the spread and relative differences between available satellite measurements, and to show whether the datasets exhibit physically consistent behaviour, as summarized and discussed in Section 5. Where possible, we provide an expert judgement on the sources of observed differences and guidance to data users regarding which datasets might be best for studies of climate variability and trends, model-measurement comparisons, and data merging activities. The water vapor comparisons presented here are part of a larger endeavour, the SPARC Data Initiative, which compares 25 different chemical tracers (among them ozone [*Tegtmeier et al.*, in preparation (a)], nitrogen oxides [*Tegtmeier et al.*, in preparation (b)]), and aerosol [*Hegglin et al.*, in preparation (b)].

2. Satellite Instruments and their Water Vapor Products

Climatologies of 13 different water vapor products from 11 limb-viewing satellite instruments (LIMS, SAGE II, UARS-MLS, HALOE, POAM III, SMR, SAGE III, MIPAS, SCIAMACHY, ACE-FTS, and Aura-MLS) (see full name of instruments in Table 1) have been compiled by the respective instrument teams to the SPARC Data Initiative for comparison. The length and vertical extent of each instrument's time series are illustrated in Figure 1. The instruments use different measurement methods (solar occultation, limb emission, or limb scattering) and operate in different wavelength bands as listed in Table 1. These measurement characteristics determine the sampling pattern (hence geographical

coverage, see detailed description for each instrument in *Toohey et al.*, [submitted]) as well as the vertical resolution of each individual instrument. The most important information on each water vapor data product are summarized in Table 2, including data version, temporal coverage, vertical extent, native vertical resolution (i.e., before interpolation onto the SPARC Data Initiative climatology grid), references to the most relevant data characterization or validation publications, and additional instrument-specific comments. It should be mentioned early on that SAGE II suffered from a shift in its retrieval channel. The exact nature of the shift and when it happened could not be established. For the V6.2 water vapor retrievals, the retrieval channel has therefore been switched from 935 nm to 945 nm in order to obtain a better match with HALOE mean values (*Thomason et. al.*, 2004). The SAGE II and HALOE mean biases versus other datasets as derived in this manuscript can therefore not be considered independent. Also, note that MIPAS operated in two different measurement modes, namely at full spectral resolution from 2002-2004 (denoted as MIPAS(1)), and after a technical problem with the interferometer at reduced spectral resolution from 2005-2010 (denoted as MIPAS(2)). Two different data products were also obtained from SMR, which are retrieved at different wavelengths and cover different altitude regions (see Table 2). More detailed information on the different instruments and how the climatologies were compiled can be found in *SPARC Data Initiative* [in preparation] and the SPARC Data Initiative overview paper, which is part of this special issue [*Hegglin et al.*, in preparation (a)].

3. Methods

We here provide a short summary of the methods used to evaluate the SPARC Data Initiative water vapor time series. More detailed information on the evaluation approach or

instrument-specific data preparation and handling can be found in *SPARC Data Initiative* [in preparation]. Note that within the SPARC Data Initiative, agreement between instruments is defined using the following terminology: excellent agreement (up to ±2.5%), very good agreement (up to ±5%), good agreement (up to ±10%), reasonably good agreement (up to ±20%), considerable disagreement (up to ±50%), and large disagreement (up to ±100%). All these numbers are with respect to the multi-instrument mean (MIM), so that where two instruments show excellent agreement of ±2.5%, the absolute difference between them is 5%.

3.1. Water Vapor Climatology Construction and Uncertainty

Zonal monthly mean time series of water vapor (in volume mixing ratio, *VMR*) have been calculated for each instrument on the SPARC Data Initiative climatology grid, using 5° latitude bins (with mid-points at 87.5°S, 82.5°S, 77.5°S, ..., 87.5°N) and 28 pressure levels (300, 250, 200, 170, 150, 130, 115, 100, 90, 80, 70, 50, 30, 20, 15, 10, 7, 5, 3, 2, 1.5, 1, 0.7, 0.5, 0.3, 0.2, 0.15, and 0.1 hPa). To this end, profile data have been carefully screened before binning and a hybrid log-linear interpolation in the vertical has been performed (except for ACE-FTS which bins in log-pressure, see *Jones et al.* [2012]). For instruments that provide data on an altitude grid, a conversion from altitude to pressure levels is performed using retrieved temperature/pressure profiles or meteorological analyses (ECMWF, GEOS-5, or NCEP). Similarly, this information is used to convert retrieved number densities into *VMR*, where needed. Along with the monthly zonal mean value, the standard deviation and the number of averaged data values are given for each grid point.

Interpretation of the differences between the individual water vapor climatologies will need to take into account several sources of uncertainty, including systematic errors of both the measurements and the climatology construction. Random measurement errors have little impact on the climatological means, however measurement biases (e.g., related to retrieval errors) will introduce systematic differences between an individual instrument's climatology and the truth. Differences in the climatologies from the truth arise also from sampling biases [*Toohey et al.*, submitted] and differences in the averaging technique used to produce the climatologies [*Funke and von Clarmann*, 2012]. Since the overall uncertainty of the climatology is not accessible in a consistent way from bottom up estimates for all of these datasets, we use here as an approximate measure of the uncertainty in each monthly mean climatology the standard error of the mean (*SEM*):

$$SEM = \sigma/\sqrt{n}, \tag{1}$$

where σ is the standard deviation of the measurements and n the number of measurements at each grid point. The range of twice the *SEM* can be loosely interpreted as the 95% confidence interval of the monthly mean. Although sampling patterns and densities differ greatly between different instruments, the *SEM* has been shown to generally produce a conservative estimate of the true random error in the mean for both solar occultation and dense sampling patterns [*Toohey and von Clarmann*, 2013]. This is due to the fact that sampling by satellite instruments is generally roughly uniform with respect to longitude. It should be noted however that the *SEM* does not contain the influence of irregular or incomplete sampling of the month and latitude band, which can produce sampling biases in the climatologies [*Toohey et al.*, submitted].

3.2. Evaluation diagnostics

A set of standard diagnostics is used to investigate and test the differences between the water vapor time series obtained from the different instruments. The diagnostics include annual and monthly zonal mean climatologies, vertical and meridional mean profiles, and seasonal cycles for a single year or averaged over multiple years. In addition, evaluations of inter-annual variability, the tropical tape recorder, and polar vortex dehydration, which test the physical consistency of the datasets, are carried out. The evaluation methods are briefly described in the following. The definition of different altitude regions in the atmosphere as used throughout the study is given in Table 3.

3.2.1. Multi-Instrument Mean Reference

In past assessments of water vapor data products [e.g., *SPARC WAVAS*, 2000], HALOE has often been used as point of reference. Here we use a different approach, namely to reference with respect to the multi-instrument mean (*MIM*). The *MIM* is calculated by taking the mean of all available instrument climatologies within a given time period of interest. Note, that the *MIM* does not represent the best estimate of the atmospheric state, since all instruments are included in its calculation regardless of their quality and without any weighting applied to them. If measurements from a particular instrument are deemed unrealistic, they are not included in the *MIM*. The relative percentage differences between the water vapor mixing ratios of an instrument ($\chi_{instrument}$) and the *MIM* (χ_{MIM}) are then given by:

$$100 * (\chi_{instrument} - \chi_{MIM})/\chi_{MIM}. \quad (2)$$

One always has to keep in mind when interpreting relative differences with respect to the *MIM* that the composition of instruments from which the *MIM* was calculated may

have changed between time periods. Hence, changes in derived differences are not to be interpreted as changes in the performance (or drifts) of an individual instrument. Also, if there is an unphysical behaviour in one instrument, the *MIM* and thus the differences with respect to the *MIM* of the other instruments will most certainly reflect this unphysical behaviour as well. Finally, if one instrument does not have global coverage for every month some sampling biases may be introduced into the *MIM*.

3.2.2. Cross-Sections and Profiles

Since water vapor is considered a long-lived species in the stratosphere, we use single- or multi-year annual averages of the monthly zonal mean cross-sections for comparison. Multi-year annual averages are preferred since they limit potential sampling errors due to inter-annual variability, e.g. through the Quasi-Biennial Oscillation (QBO). Exceptions to this are the single-month comparisons of LIMS due to the shorter lifetime of this mission (<1 year) and UARS-MLS due to the strongly varying aerosol loading of the atmosphere during its lifetime caused by the Mt. Pinatubo eruption, which may aggravate sampling biases in multi-annual means. Also note that sampling biases will have a larger impact on the multi-annual mean evaluations closer to the tropical tropopause, where natural variability is known to be large. In addition, vertical and meridional profile comparisons are hence performed on a monthly basis in order to provide information on the monthly behavior and check on the impact of sampling issues on the instrument differences.

3.2.3. Seasonal Cycle Analysis

For the comparison of seasonal cycles, a multi-year approach has been chosen as well. Note that a reliable estimate of the mean annual cycle would require averaging the data over at least one or two QBO-cycles (approximately 3 or 5 years) in order to properly ac-

count for the strong QBO-dependence water vapor exhibits (e.g., *Fueglistaler and Haynes* [2005]). The plots of the seasonal cycles include the *MIM* together with its standard deviation, which is a measure of the range of mean values obtained by the different instruments. A combined annual and semi-annual fit has been applied to all the available monthly mean values of a single instrument in order to yield a seasonal cycle that is comparable for all instruments, including those that do not measure all months of the year. In addition to the seasonal cycle plots, Taylor diagrams [*Taylor*, 2001] are shown in order to compare the different instruments in a more quantitative way. Taylor diagrams offer a visual summary of how well the seasonal cycle of a certain instrument represents the seasonal cycle of a reference field, in our case the *MIM*. Three measures of agreement are shown on Taylor plots (*Taylor*, 2001; see also *Hegglin et al.* [2010] for an illustrative example): the correlation on the azimuthal axis, which represents how well the phase of the seasonal cycle is measured by the instrument; the normalized amplitude on the radial axes, which indicates how well the strength of the seasonal cycle is captured; and the skill factor, given with the light grey lines, which is a combination of the other two measures and therefore summarizes the overall performance of an instrument.

3.2.4. Time Series of Absolute and Deseasonalized Values

Time series of absolute values and deseasonalized anomalies are used to analyze intra-annual and inter-annual variability in the water vapor datasets.

Time series based on absolute values in time-pressure coordinates are used to evaluate the tropical tape recorder and polar dehydration as they are represented in each instrument's climatologies. For the tropical tape recorder, we average vertical profiles between 15°S and 15°N as chosen in the original study by *Mote et al.* [1996] in order to get a strong

signal unaffected by the subtropical mixing zones. For the polar vortex dehydration, we use profiles between 60°S and 90°S. Only the dehydration in the southern polar vortex is examined since the northern polar vortex is not as strongly isolated.

In addition, time series of deseasonalized anomalies are shown for selected latitude bands and pressure levels. Monthly anomalies are calculated for each instrument by subtracting the multi-year monthly mean from its monthly mean values. The multi-year monthly mean is then averaged over all years of the whole time series available from each instrument. No additional adjustments are applied to the anomaly time series.

3.2.5. Summary Evaluations

We use two different sets of summary plots in order to present an overview of the findings in Section 5. The first set of summary plots highlights the uncertainty estimate in our knowledge of the atmospheric mean state. It shows the annual zonal mean *MIM*, minimum (*MIN*) and maximum (*MAX*) fields. The latter two are the minimum and maximum values found at each grid point across all instruments. In addition, the difference between *MAX* and *MIN* as well as the standard deviation over all instruments are presented in absolute and relative values to give information on the spread around the *MIM*.

The second set of summary plot highlights specific inter-instrument differences in selected regions of the atmosphere, emphasizing those datasets that are consistent with one another and those that are not. The differences are plotted for each instrument and region in form of the median (or mean) deviation from the MIM, calculated over all the differences from the MIM at each individual grid point within the selected region. The regions are divided into different altitude ranges (300-100 hPa; 100-30 hPa; 30-5 hPa; 5-1 hPa; 1-0.1 hPa) and into the extra-tropics (40°-80°S and N) and the tropics (20°S-20°N).

In addition, the median absolute deviation (MAD) is provided for each instrument and region. The MAD over the sample $x = (x_1, \ldots, x_n)$ is defined as:

$$MAD = median(|x - median(x)|) \quad (3)$$

and represents the interval around the median that contains 50% of the data [*Rousseeuw and Croux*, 1993]. For comparison, the range indicating the mean $\pm$ 1σ is also indicated. Instruments that were not measuring during the reference time period were added to the plot using instruments that measured during both periods as transfer standards. This procedure also ensures that potential trends in water vapor do not influence the comparison. However, if the instrument used as the transfer standard exhibits a drift between the two time periods, the comparison would be biased by the magnitude of the drift.

4. Results and Discussion

4.1. Cross-section Comparisons

4.1.1. LIMS (1978-1979) versus SAGE II (1984-1990)

Figure 2 shows a comparison of monthly zonal mean water vapor fields between LIMS and SAGE II. Note that the MIM is not weighted by the length of the observation periods, such that SAGE II has the same weight as LIMS despite having more years of observations. Also, as mentioned in Section 2, the comparison with SAGE II V6.2 water vapor mean values is affected by the one-off adjustment of SAGE II to match with HALOE in the mid-1980s. The figure reveals the key features of the water vapor distribution in the middle atmosphere, which is the result of a combination of the Brewer-Dobson circulation

and a stratospheric source of water vapor. Air entering the stratosphere is dehydrated as it passes through the very cold tropical tropopause, creating a minimum in water vapor just above the tropopause. As the air ascends to higher altitudes, it gains water vapor through the oxidation of methane [*Bates and Nicolet*, 1950]. Isentropic mixing between the ascending branch of the Brewer-Dobson circulation in the tropics (with low water vapor values) and the descending branch in the extra-tropics (with high water vapor values) then produces typical downward and poleward sloping tracer isopleths. Dehydration and subsequent sedimentation of ice particles in the cold winter polar vortex can lead to an additional minimum in the lower stratosphere at high latitudes [*Kelly et al.*, 1989].

The comparison reveals quantitatively that LIMS and SAGE II show very good to excellent agreement in the tropics (within ±2.5-5% of the *MIM*, corresponding to inter-instrument differences of 5-10%) and mostly agree well in the extra-tropics (within ±5-10% of the *MIM*, or 10-20% inter-instrument differences), although it should be kept in mind that the measurements from the two instruments do not overlap in time. Generally, SAGE II values are somewhat lower (higher) than LIMS values below (above) 10 hPa. LIMS exhibits rather atypical isopleths (also when compared to other, later instruments) that with the lowest tropical values not reaching far enough into mid-latitudes. As a consequence, the differences from the *MIM* increase moving to higher latitudes. Validation of LIMS water vapor V6.0 with a limited number of available correlative profile measurements at mid-latitudes, confirm that LIMS between 10 and 70 hPa is higher by about 10-15% although within the stated measurement uncertainties of the respective instruments *[Remsberg et al.*, 2009]. Below 80-100 hPa, the differences from the *MIM* increase to over ±20% across all latitudes, with SAGE II showing negative and LIMS showing pos-

itive deviations. The disparity in vertical resolution of the two instruments (see *Table* 2) is likely the reason for the large differences closer to the tropopause, where strong vertical gradients in water vapor are found.

4.1.2. SAGE II, HALOE, and UARS-MLS (1992)

Figure 3 shows cross-sections of the relative differences in the monthly zonal mean water vapor of SAGE II, UARS-MLS, and HALOE with respect to their *MIM*. Note that the year 1992 is not ideal for comparison due to a heavier aerosol loading in the stratosphere after the Mount Pinatubo eruption, which may adversely affect the retrievals of solar occultation measurements. The inter-instrument differences derived from this time period may therefore not be consistent with differences derived for later time periods. However, it is the only direct comparison we have with the measurements from the UARS-MLS instrument.

The relative differences from the *MIM* are relatively small with values between ±2.5% and ±5% throughout most of the MS, US and LM indicating excellent to very good agreement between the instruments. HALOE values generally lie between the lower UARS-MLS values and the higher SAGE II values. Indeed, *Pumphrey* [1999] showed that the UARS-MLS water vapor data version used here (called prototype version 0104 at that time) yielded values uniformly smaller than HALOE (by 0.1 to 0.4 ppmv) and about 0.6 ppmv smaller than the ATMOS results obtained from the Space Shuttle, but compared well to the average of 16 coincident frost point hygrometer profiles. Note that averaging over all available years of UARS-MLS (1991-1993) renders the comparison less noisy (not shown), with UARS-MLS being consistently more negative than HALOE, except in the

Antarctic LS, where HALOE exhibits a strong low bias compared to both SAGE II and UARS-MLS.

4.1.3. SAGE II, HALOE, POAM III, SMR, SAGE III, MIPAS, SCIAMACHY, ACE-FTS, and Aura-MLS (1998-2008)

Figures 4 and 5 show the annual zonal mean and relative difference cross-sections, respectively, for climatologies obtained over the years 1998-2008. Despite the fact that the climatologies of the individual instruments span different time periods (as indicated in the figure titles), this approach has been chosen so that a maximum number of instruments can be compared, and to limit the influence of reduced sampling by HALOE and SAGE II in the early 2000s. The results for the 1998-2008 time period are consistent with results obtained from single-year evaluations for 2003 or multi-year averages spanning 2006-2009 (not shown), providing evidence that trends in water vapor over this time period have only a minor impact on the comparison. Note that the evaluation of the 1998-2008 climatologies will be used as the basis for the summary plots in Section 5.

Most of the instruments capture the main features of the water vapor distribution well, including the downward sloping isopleths towards higher latitudes, the minima above the tropical tropopause and within the polar vortex region in the Southern Hemisphere (SH; which stems from dehydration within the polar vortex during winter and is strong enough to affect the annual mean), and the maximum in the water vapor distribution in the lower mesosphere (Figure 4). However, large differences in the absolute values can be seen in the tropical UT, where HALOE VMRs are notably smaller than seen in the *MIM*, and where SMR(2) VMRs show much flatter water vapor isopleths than the other instruments.

Figure 5 shows that overall best agreement is found in the MS. The older set of instruments (HALOE and SAGE II, with SAGE II however not being fully independent from HALOE as discussed above), and also SMR(1) above 10 hPa, exhibit much lower values than the newer set of instruments (MIPAS(1), MIPAS(2), SAGE III, ACE-FTS, and Aura-MLS) throughout most of the stratosphere with differences from the *MIM* of up to -10%. Note that a validation study by *Nedoluha et al.* [2007] in the mesosphere using Water Vapor Millimeter-wave Spectrometer (WVMS) measurements over Mauna Loa also indicates that HALOE is biased low 10%, while ACE-FTS and Aura-MLS are within ±0.5-1.5%. On the other hand, POAM III exhibits rather large positive deviations from the *MIM*. In the USLM, SMR(1) shows the largest negative differences (around -15%) and Aura-MLS the largest positive differences from the *MIM* (around +10%). SMR(1), however, agrees very well with the other instruments between 50 and 10 hPa. MIPAS(2) and MIPAS(1) report positive deviations compared to the *MIM* through most of the stratosphere. Below 100 hPa, SAGE II and HALOE show deviations from the *MIM* that are larger than -20%. In contrast to MIPAS(1), MIPAS(2) shows negative differences in the LM and positive differences in the UTLS (except in the tropical UT). The low bias in MIPAS(2) is explained by a neglect of non-LTE effects in the retrieval of the data (see *Stiller et al.* [2012] for full explanation). Aura-MLS shows a sandwich-like layer of negative deviations around 200 hPa in between regions of positive deviations; it is known that the current MLS retrieval scheme tends to underestimate the low water vapor values near the hygropause [*Read et al.*, 2007; *Voemel et al.*, 2007]. SMR(2) exhibits relative differences below 100 hPa of up to +100%, and above 100 hPa up to -40% in the extra-tropics and -20% in the tropics.

4.2. Vertical and Meridional Profile Comparisons

Monthly mean vertical and meridional water vapor profiles and their relative differences to the *MIM* are shown for 1978-1990, 1991-1993, and 1998-2008 in Figures 6 and 7, respectively. In addition to the relative differences, the *SEM* for each profile is shown, as an approximate measure of the random error in each climatology. Results of the vertical profile comparisons are shown for a SH tropical or subtropical latitude band and a NH mid to high latitude region (chosen to maximize the number of available instruments). Results for the meridional profiles are shown for the 10 and 80 hPa levels (chosen since they reflect typical regions of scientific interest, i.e., the water vapor entry level in the tropics).

For the 1978-1990 time period, Figure 6 highlights again the excellent to very good agreement (within 5-10%) between LIMS and SAGE II in both the tropics and extra-tropics above around 70 hPa, with very small *SEM* values for both instruments. Below 70 hPa, the deviations from the *MIM* increase substantially with maximum disagreement in the tropopause region. Similar behavior is seen in the meridional profiles (see Figure 7), which also reveals that the deviations from the *MIM* at 80 hPa are greater in the tropics than in the extra-tropics. This may be due to residual emissions from cloud tops in the tropics, since this version of LIMS water vapor data has not been thoroughly screened for clouds.

For the 1991-1993 time period, the vertical and meridional profiles emphasize the very good to excellent agreement between UARS-MLS, SAGE II and HALOE. In comparison with the cross-section evaluation for 1992 (Section 4.1.2), the profile evaluation based on the longer time period 1991-1993 now show that, as a result of averaging over a longer time

period, UARS-MLS is consistently lower than HALOE (and SAGE II, in regions where HALOE and SAGE II do not agree), with small *SEM* values indicating a significant difference in the climatologies of these instruments. SAGE II, on the other hand, shows a somewhat less defined (or noisier) climatological profile (with larger *SEM* values). As expected from the adjustment of the SAGE II retrieval channel to yield a better match with HALOE, HALOE and SAGE II agree within their uncertainties throughout most of the LS and MS, at least at higher latitudes. However, the instruments disagree in the tropics. Also, in the region below 100 hPa in both the tropics and extra-tropics, their deviations from each other increase substantially. This result indicates that SAGE II has the potential to add information to the HALOE measurements.

For the 1998-2008 period, the monthly comparisons generally confirm the findings of the annual zonal mean cross-section evaluations, but they indicate somewhat larger inter-instrument differences. Most instruments agree well within ±5% between 100 and 10 hPa, with increasing differences above and below. Exceptions are SAGE II, which shows much lower values in the LS and MS, and POAM III, which shows much higher values than the other instruments (up to +20% from the *MIM*). SMR(1) is also on the low side of the other climatologies. At 80 hPa (Figure 7), the spread in the climatologies increases strongly to ±20% in the tropics, with somewhat smaller discrepancies in the extra-tropics. SMR(2), SAGE II and to a somewhat lesser extent HALOE are all on the low side of the *MIM*. SCIAMACHY shows a large positive deviation from the *MIM* of up to 40% in the tropical region during May, but agrees well with the other instruments in the extra-tropics. MIPAS(1), MIPAS(2), ACE-FTS and Aura-MLS agree within 15%. The *SEM* values are generally much smaller for the limb emission than the solar occultation sounders and

are larger in the UTLS than in the MS. The relative differences between the individual instruments in the UTLS are therefore less well defined.

4.3. Seasonal Cycles

Water vapor exhibits strong seasonal cycles in both the tropical and extra-tropical UTLS due to its dependence on transport and Lagrangian cold-point temperatures [*Fueglistaler et al.*, 2009; *Hoor et al.*, 2010]. Most attention has focused on the tropics between 80 and 100 hPa, where the stratospheric entry value of water vapor is slaved to the seasonally changing cold-point temperatures (e.g., *Mote et al.* [1996]; *Fueglistaler and Haynes* [2005]; *Fujiwara et al.* [2010]). However, the seasonal cycle is also of interest in the extra-tropics (especially at levels lower than 100 hPa) where it reflects the impact of stratosphere-troposphere exchange, and hence allows for the evaluation of transport processes in chemistry-climate models [Hegglin et al., 2010].

Figure 8 shows the seasonal cycles in water vapor and corresponding Taylor diagrams at 70, 100, and 170 hPa in the tropical lower stratosphere averaged over the years 1998-2008 for all available instruments. Following the known seasonal cycle in the cold-point tropopause temperature, water vapor mixing ratios show a minimum in NH spring and a maximum in NH autumn. Best agreement with the *MIM* in terms of absolute mean values is found at 70 hPa (with a 1σ inter-instrument spread of 12%). The inter-instrument spread increases to 22.5% at 100 hPa, and 30% at 170 hPa. SCIAMACHY shows the highest mean values at all levels, with largest differences at 100 hPa. This is partially explained by the fact that the native retrieval levels of SCIAMACHY lie relatively far above and below 100 hPa (at 70 and 130 hPa), which leads to smearing of the real values across the region of strong gradients in water vapor at the tropopause. HALOE and

SAGE II are both much lower than the *MIM*. The few tropical monthly mean data points produced by ACE-FTS are distributed such that the amplitude and phase of the fitted seasonal cycle agree well with other instruments.

We find best agreement between the monthly mean values of ACE-FTS, Aura-MLS, MIPAS(1) and MIPAS(2) at 70 hPa, with SMR(2) joining this group at 100 hPa. However, Taylor diagrams reveal better agreement between the instruments' seasonal cycle in terms of correlation and amplitude of the seasonal cycle at 100 hPa. MIPAS(1) and MIPAS(2) are well correlated with the *MIM* at these two levels, however show amplitudes that are (about half) too low compared to the *MIM*. These low amplitudes are explained by state-dependent averaging kernels, causing MIPAS to resolve atmospheric structures better in a more humid atmosphere, which leads to a reduced amplitude of the seasonal cycle. Application of the seasonally varying averaging kernels from MIPAS to the other instruments (using the same principle as when validating satellite observations against ozonesonde data) would hence improve the results for these comparisons. At 170 hPa, the SAGE II, HALOE and MIPAS(2) data agree best in phase (correlation of 0.8) and amplitude as can be seen in the corresponding Taylor diagram. However, MIPAS(1) and Aura-MLS show much too small and SCIAMACHY and ACE-FTS much too high amplitudes. Moreover, Aura-MLS does not capture the phase correctly.

Seasonal cycles in water vapor for the Southern and Northern Hemisphere (NH) mid to high latitudes at different pressure levels (100, 150, and 250 hPa) are displayed in Figure 9. The maximum in the seasonal cycle at 100 hPa in these regions is expected to be slaved to the tropics with a few months delay. However, at lower altitudes the maximum is shifted to summer, since a higher tropopause during the minima in the seasonal downwelling of

the Brewer-Dobson circulation allows for more tropospheric in-mixing. We find that the mean values of the seasonal cycle in both hemispheres are better constrained at 100 and at 150 hPa (with a 1σ-spread of 15-20%) than at 250 hPa where the spread increases to 25-50% during summer peak values. In the SH, the climatologies are less consistent between instruments at the 100 and 150 hPa levels, possibly due to the fact that the seasonality here is strongly influenced by in-mixing of dehydrated polar air masses following the winter months. SAGE III here is an outlier showing too large amplitudes at 150 and 250 hPa. In the NH, the instruments perform best at 150 hPa with the exception of SAGE III and Aura-MLS. This is readily seen in the Taylor diagram, where these two instruments exhibit a skill factor of 0.7 only, while all the other instruments exhibit skill factors of higher than 0.9. Aura-MLS shows a phase shift in the seasonal cycle at 150 hPa, while it performs well in both amplitude and phase at 100 and 250 hPa. SAGE II, while showing excellent agreement with the *MIM* at 100 and 150 hPa, is an outlier at 250 hPa with a maximum that peaks too late. MIPAS(2) shows the most robust results performing very well at all altitudes and in both hemispheres. HALOE is on the low side of the *MIM* and its data should not be used at 250 hPa.

The difficulties of representing the annual cycle in water vapor at different levels in the UTLS are related to the strong vertical gradients in water vapor found across the tropopause and the narrow vertical region over which the annual cycle extends. Both factors require high vertical resolution measurements and/or high vertical sampling to be adequately resolved. Also, instrumental limitations result from cloud interference and high extinction exist in this altitude region. As a consequence it is generally found that instruments with either less frequent sampling or low vertical resolution show less robust

results in the different regions. UTLS-specific evaluations using tropopause coordinates or equivalent latitude may help to improve the comparisons in the future.

4.4. Inter-annual Variability

In addition to the pronounced seasonal cycle, water vapor is characterised by non-seasonal variations related to ENSO (the El Nino Southern Oscillation) and the QBO [e.g., Niwano et al., 2003; Randel et al., 2004], and to a smaller extent by inter-annual variability in tropical convection or polar vortex temperatures. Long-term water vapor variability involves changes in methane (the stratospheric source gas for water vapor) and decadal scale climate variability. The evaluation of inter-annual variability using deseasonalized anomalies yields insight into whether an instrument can produce physically consistent time series in comparison to other datasets. While the longer-term evolution of the anomalies is indicative of the long-term stability of the instruments, some monthly differences are likely to be introduced by noise or sampling issues.

Figure 10 shows time series of deseasonalized water vapor anomalies at 10 and 70 hPa in the tropics, and at 10 and 80 hPa in the NH extra-tropics between 1997 and 2010. The evaluation starts in 1997 in order to avoid the years affected by the 1991 Mt. Pinatubo eruption. The different instruments show very good agreement with generally consistent long-term tendencies, and with the QBO leaving the most pronounced signature in the anomalies. Note that while the QBO is a tropical phenomenon, it has also a distinct influence on extra-tropical water vapor, although with an attenuated signal at 80 hPa due to mixing processes, and a delay compared to the tropical signal related to stratospheric transport time scales. It is noteworthy that the instruments also agree on the breakdown

of the QBO signal on the tropical 10 and 70 hPa and the extra-tropical 80 hPa levels in the early 2000s and after 2008.

In the tropics at 10 hPa, HALOE and SAGE II show somewhat larger intra- and inter-annual variations than SMR(1), MIPAS(1), MIPAS(2), Aura-MLS, which is possibly related to the limited sampling patterns of the occultation instruments. ACE-FTS on the other hand, despite limited tropical sampling, agrees well with the rest of the instruments. At 70 hPa, inter-annual variability and also month-to-month fluctuations are stronger than at 10 hPa. The spread in the instruments is larger, with SMR(2) showing spike-like structures after 2006, which are not seen in the other instruments. MIPAS(2) shows a somewhat smaller amplitude in its inter-annual variations than the other instruments.

In the extra-tropics at 10 hPa, all instruments (now including POAM III and SAGE III, which did not measure in the tropics) seem to follow month-to-month fluctuations extremely well. SAGE III starts slightly at too positive anomalies or ends at too negative anomalies indicating a potential sampling issue in the instrument, which needs further investigation. SMR(1) exhibits a spike at the end of 2010 that is not seen in the other instruments. At 80 hPa, POAM III agrees well with HALOE and SAGE II in the early years, however shows somewhat larger negative fluctuations in the monthly anomalies between 2003 and 2005 compared to the other instruments. SAGE III again shows a potential drift, which seems somewhat stronger than at 10 hPa. SMR(2) is very noisy at this level. ACE-FTS, Aura-MLS and MIPAS(2) show the best agreement.

4.5. Tropical Tape Recorder

The atmospheric tape recorder [*Mote et al.*, 1996] is one of the most pronounced spatio-temporal patterns in equatorial water vapor, showing the slow upward propagation of a

minimum in water vapor from the tropical tropopause region up to altitudes of around 30 km. The signal is produced by seasonal variations in tropical tropopause temperatures, which determine the water vapor saturation mixing ratios in air masses entering the tropical stratosphere. A realistic characterization of the tape recorder is a key aspect of the physical consistency of the different datasets, provided that the sampling is adequate. No tape recorder could be produced for SAGE III and POAM III, which have no tropical coverage.

Figures 11 and 12 show the tape recorder of the *MIM* and of each individual instrument, and the relative differences to the *MIM* for each instrument, respectively. In general, the instruments reproduce the main features of the tape recorder well, with a strong minimum in water vapor values during NH winter, which slowly propagates upward over time. Note in this evaluation MIPAS(1) and MIPAS(2) are shown in the same panel, since they cover different time periods. ACE-FTS has only limited temporal coverage in the tropics and hence interpolated data in time and altitude are shown. Despite this limitation, the interpolated ACE-FTS data show, a realistic tape recorder, and the relative differences to the *MIM* seem within the range of the other instrument differences. Although a tape recorder is also visible for SCIAMACHY, the water vapor minimum just above the tropical tropopause is much weaker, and the higher mixing ratios reach further into the stratosphere as seen for the *MIM*. As discussed earlier, this is due to the coarse sampling of SCIAMACHY in the tropopause region that leads to strong smearing of the values across the tropopause. SMR(2) shows much lower values throughout the LS than the other instruments.

The evaluation of the differences between the climatologies and the *MIM* (Figure 12) reveals that for the period 2000-2005, SAGE II and HALOE agree well, with differences that have a rather noisy structure, implying that the two instruments have no systematic biases and that the tape recorder signal they reproduce is physically consistent. Both instruments, however, measure lower values than the newer generation of instruments (SMR(1) and MIPAS(1)) that contribute to the *MIM* at the beginning of 2002. Since SMR(1) yields the most negative deviations from the *MIM* after 2004 when more instruments come on-line, it follows that HALOE (and SAGE II for this matter) would likely be on the low side of these as well. In the later period, MIPAS(2), SCIAMACHY, and Aura-MLS exhibit structures in the LS that resemble the tape recorder itself, implying a systematic difference, which may be due to the effects of different vertical resolutions (see Table 3). Resolution issues would affect the derived amplitude of the tape recorder, which is often used as a diagnostic in model-measurement comparisons. In addition, MIPAS(2) and Aura-MLS have higher values in the MS when compared to ACE-FTS and SMR(1). SMR(2) shows negative deviations of >20% from the *MIM* in the 50-100 hPa range. However, the noise in the relative deviations indicates that it captures the seasonal cycle reasonably well compared to the other instruments.

A tape recorder has also been derived for the LIMS instrument (not shown; cf. *SPARC Data Initiative* [in preparation]). While the tape recorder shows a distinct minimum in water vapor above the tropical tropopause, the signal does not seem to propagate into the middle stratosphere. Note that the data are very limited in time (only five months of LIMS data are available). A tape recorder has also been found for UARS-MLS in the LS and MS as demonstrated before by *Pumphrey* [1999].

4.6. Polar Vortex Dehydration

Another spatio-temporal pattern that is seen in water vapor is the descent of aged and water vapor-enriched air masses and subsequent dehydration through ice nucleation and sedimentation in the polar vortex of the SH. Since this phenomenon happens predominantly in winter/early spring, occultation instruments will not capture its full extent. However, for satellite instruments that are measuring in darkness, the evaluation provides a stringent test of whether the retrieval in this region is hampered by the presence of ice particles (although this is not a problem for instruments that measure in the microwave/sub-mm range, see Table 1). The time period 2002-2010 has been chosen, since it encompasses most of the satellite instruments used in this study and allows for the evaluation of inter-annual variability in this region.

Figures 13 and 14 shows polar vortex dehydration as seen in the *MIM* and for each instrument, and each instrument's relative differences to the *MIM*, respectively, between 2002 and 2010. The most comprehensive results are obtained from Aura-MLS and MIPAS, two emission sounders, which are able to measure water vapor during polar night. Note that in this evaluation MIPAS(1) and MIPAS(2) are again shown in the same panel. Many of the solar occultation results exhibit the correct physical structure, however, their less frequent sampling limits the overall picture. SAGE II and HALOE measurements are lower than the MIM throughout the UTLS and MS, while ACE-FTS shows positive deviations from the *MIM* in the UT and somewhat lower values in the regions of dehydration. Note that POAM III exhibits a more uniform temporal sampling of the polar region than other solar occultation instruments. Nevertheless, POAM III shows larger positive deviations from the *MIM* than SAGE II and HALOE, which is possibly due to POAM III

sampling mostly the outside portion of the polar vortex. SMR(2) shows much lower values than the other instruments and very prominent dehydration structures that extend into the January-April period. SCIAMACHY shows consistent features, but underestimates the strength of dehydration with respect to the *MIM*. This is most probably due to the fact that only measurements at SZAs smaller than 85° were used to construct the SCIAMACHY water vapor climatologies, limiting its sampling to the outer parts of the polar vortex. Aura-MLS shows the largest dehydration within the polar vortex, and relatively strong negative deviations from the *MIM* around 200 hPa, but agrees well with POAM III at these levels. SMR(1) (not shown) performs well at higher altitudes, although it exhibits slightly lower mixing ratios than MIPAS and Aura-MLS.

5. Summary and Conclusions

The SPARC Data Initiative has performed a comprehensive comparison of 13 available water vapor products from 11 different available limb-viewing satellite instruments (LIMS, SAGE II, UARS-MLS, HALOE, POAM III, SMR, SAGE III, MIPAS, SCIAMACHY, ACE-FTS, and Aura-MLS) between 1978-2010. Note that there are more water vapor measurements available (e.g., those from SAMS [*Jones et al.*, 1986], ISAMS [*Taylor et al.*, 1993], CLAES [*Roche et al.*, 1993], or ATMOS [*Gunson et al.*, 1996]) but they could not be included in our evaluations due to a lack of resources needed to generate the zonal monthly mean climatologies. The instrument comparisons were based on a climatological approach using monthly zonal mean time series rather than the classical validation method using coincident profile measurements. Diagnostics include single- or multi-year, monthly or annual zonal mean cross-sections, inter-annual variability, seasonal cycles, and time evolution of specific physical features in water vapor such as the tropical tape

recorder or polar vortex dehydration. The following results on the atmospheric mean state, performance by region, and performance of individual instruments have been found as summarized in Figures 15 and 16.

5.1. Atmospheric mean state

Our knowledge of the atmospheric mean state in water vapor derived from the full set of instruments available between 1998 and 2008 (excluding SMR(2) and MIPAS(1) in order not to 'double-count' instruments) is best in the lower and middle stratosphere of the tropics and mid-latitudes, with a relative uncertainty of ± 2-6% (1σ) with respect to the *MIM* as can be seen in Figure 15. This uncertainty in water vapor increases toward the polar latitudes (±10% and ±15% for NH and SH, respectively), the lower mesosphere (± 15%) and the troposphere (± 30-50%). Note that the uncertainty in water vapor is largest in the subtropical jet region (30-50°N/S), which may be partly due to the large dynamical variability observed in tropopause height, which affects the climatologies due to sampling issues. The minimum in the multi-annual zonal mean of water vapor found just above the tropical tropopause shows values ranging from approximately 2.5 to 4.5 ppmv when including all instruments, with a mean of 3.5 ± 0.5 ppmv (or ±14%, 1σ uncertainty). The 1σ uncertainty is somewhat larger (±15-20%) when looking at individual months (see seasonal cycle evaluation Figures 8 and 9). The maximum found in the annual zonal mean of water vapor in the LM shows an absolute range of approximately 5.5-7.5 ppmv, with a mean of 6.5 ±0.7 ppmv (or ±9%, 1σ uncertainty).

5.2. Performance by Region

The following results are illustrated in Figure 16.

Lower Mesosphere (0.1-1 hPa): In the tropical and extra-tropical LM, the instruments agree well, within approximately 10% of the *MIM* (corresponding to inter-instrument differences of up to 20%). The more recent instruments (ACE-FTS, Aura-MLS, and MIPAS(1) and (2)) even show excellent agreement of within 5% of each other. A clear exception to this is SMR(1), which shows deviations from the *MIM* of up to -18%. Together with HALOE and UARS-MLS, SMR(1) is on the low side of the *MIM*. Earlier results from validation studies using coincident measurements from other independent instruments support these findings: UARS-MLS was found to have a low bias of 5% when compared to the ATMOS instrument (and HALOE) [*Pumphrey*, 1999]. Note that the spatial variability of the deviations within one region is relatively small for most instruments, indicated by small *MADs* (around ±3%). POAM III shows a larger range, indicated by a larger *MAD* (±6%).

Upper Stratosphere (1-5 hPa): In the tropical and extra-tropical US, the instruments show good agreement, within ±10% of the *MIM*, and very small MADs (±1.5%) for most instruments indicating a narrow distribution of deviations from the *MIM* within these regions. This means that while individual instruments may disagree with each other, their differences are well characterized. Most instruments agree even very well, within ±5%. Exceptions in the tropical region are UARS-MLS and SMR(1), which show larger negative deviations, and MIPAS (2), which shows a larger positive deviation from the *MIM* than the other instruments. Exceptions in the extra-tropical regions are LIMS, SMR(1), and UARS-MLS. POAM III data in the extra-tropics show the highest values, although close to those from MIPAS (1) and MIPAS (2).

Middle Stratosphere (5-30 hPa): In both the tropical and extra-tropical MS, most instruments agree very well to within ±5% of the *MIM*. Notable is the excellent agreement (within ±2.5%) between ACE-FTS, Aura-MLS, HALOE, LIMS, MIPAS (1) and MIPAS(2) in the extra-tropics. Small *MADs* (mostly ±3 to ±4%) indicate small variability in the deviations and hence that the instrument differences are well characterized. Exceptions in the tropics are UARS-MLS, LIMS, and SMR(1), in the extra-tropics UARS-MLS, SCIAMACHY, and POAM III.

Lower Stratosphere (30-100 hPa): In the tropical LS, the instruments show only reasonably good agreement, mostly within ±20% of the *MIM*. The agreement is much better in the extra-tropical LS with deviations of only ±5% of the *MIM*. Exceptions are LIMS, POAM III and UARS-MLS with deviations of ±10% of the *MIM*, and SMR(2) with a deviation of -22% from the *MIM*. Very good agreement is found for the ACE-FTS, Aura-MLS, HALOE, MIPAS(1), MIPAS(2), SAGE II, SAGE III, SCIAMACHY, and SMR(1). The *MADs* indicate better defined deviations in the extra-tropics than in the tropics.

Upper Troposphere/Lower Stratosphere (100-300 hPa): Considerable disagreement between the instruments is found for the lowest levels between 100 and 300 hPa of both the tropical and extra-tropical UTLS, with maximum differences from the *MIM* of ±40% and ±30%, respectively. Nevertheless, very good agreement (within ±5% of the *MIM*) is found for Aura-MLS, MIPAS(1), MIPAS(2), POAM III, and SAGE III in the extra-tropics. Large *MADs* (±10% or more) indicate spatial inhomogeneity of the deviations in the two regions and hence poorly defined measurement behaviour. Note SMR(2) shows deviations from the *MIM* of more than +50% in this region and hence is off the

scale. The poor agreement in the UTLS may partly be explained by sampling issues and partly by the difficulties the instruments encounter in measuring accurately in the UTLS. Large dynamical variability and steep tracer gradients across the tropopause limit especially instruments with low temporal (occultation sounders) or vertical resolution (emission sounders). Also, cloud interference and saturation of the measured radiances pose retrieval challenges.

5.3. Instrument-specific Conclusions

LIMS (V6.0) provides the earliest water vapor observations available to the SPARC Data Initiative. The LIMS record extends over only a few months. Using SAGE II as a transfer, LIMS shows very good agreement (±5% of the *MIM*) in the MS and the tropical US. However, it shows large negative deviation from the *MIM* of around -12% in the extra-tropical US, and large positive deviations from the *MIM* of +15% in the LS and +30 to +40% in the UTLS (between 100 and 300 hPa), respectively.

SAGE II (V6.2) provides the longest water vapor record, however, users should be cautious about using SAGE II water vapor data for trend studies because of a shift in its retrieval channel (*Thomason et. al.*, 2004). The exact nature of the shift and when it happened could not be established and there is no indication from our results that the shift is affecting the long-term stability of the instrument. For the V6.2 retrievals, the retrieval channel has been switched from 935 nm to 945 nm, as this led to a better match with HALOE mean values. Similarities between SAGE II and HALOE mean biases versus other datasets or versus the MIM can be understood in terms of this adjustment of SAGE II to HALOE data. However, while the two datasets do not have independent mean values because of the SAGE II adjustments related to this channel drift issue, the

time-dependent changes between the two datasets can be viewed as being independent. Also, our evaluations have shown that SAGE II and HALOE exhibit somewhat different altitudinal and latitudinal structures, providing evidence that SAGE II observations have the potential to add information on atmospheric water vapor distributions. SAGE II is seen to perform very well in inter-annual variability evaluations for which its use should not be restricted. Note that SAGE II exhibits a strong (however known) bias above about 3 hPa due to which the data above this level are not provided in the SPARC Data Initiative monthly zonal mean climatologies.

HALOE (V19) is the hitherto most used water vapor dataset. Our evaluations indicate that HALOE's water vapor has a slight low bias throughout the atmosphere. Deviations from the *MIM* are found to be around -5% through most of the stratosphere and LM consistent with results from *SPARC WAVAS* [2000]. HALOE's low bias is strongest in the UTLS (between 100 and 300 hPa) to values lower than -20%, and the instrument fails to reproduce the seasonal cycles at 200 hPa and below in both the tropics and extra-tropics. However, note that aside from the low bias, HALOE resolves the amplitude and phase of the seasonal cycle and also the magnitude of inter-annual variability well at all SPARC Data Initiative levels between 100 and 170 hPa.

UARS-MLS (V6) offers water vapor measurements over a limited time period in the early 1990s. The measurements are seen to be about 5% lower than HALOE through most of the atmosphere, a result confirmed by validation with in-situ measurements.

SAGE III (V4.0) is limited to the extra-tropics, however, shows excellent agreement with the *MIM* throughout the atmosphere and even in the UTLS (between 100-300 hPa).

While its limited availability restricts its use to a small number of evaluations, its data record may be considered as a transfer for merging activities.

POAM III (V4.0) is another instrument with somewhat limited temporal and spatial coverage. The biases derived in our evaluations are consistent with earlier validation studies. POAM III is biased high throughout the stratosphere with somewhat larger deviations from the *MIM* in the SH (>20%) than in the NH (>10%). However, it performs very well (within 5% from the *MIM*) at the lowest levels (100-300 hPa). Despite the positive biases, the instrument performs well in evaluations of inter-annual variability and compares well to SAGE II and HALOE, making it a potentially useful instrument to study climate variability or to merge HALOE and SAGE II with the newer instruments.

The **SMR(2) (V2.0)** water vapor product (derived using the 544 GHz-band) does not exhibit a correct tropopause-following structure of the tracer isopleths and the values are too high below and too low above 100 hPa, respectively. Nevertheless, once its time-averaged bias is removed, SMR(2) exhibits a reasonably accurate quantification of the strength of the inter-annual variability in the tropics and also shows an amplitude and phase of the tropical seasonal cycle that agree well with the *MIM*. However, the data are less consistent in the extra-tropics. The data product needs further improvement and the recommendation is to restrict its use to between 100 and 50 hPa in the tropics. **SMR(1) (V2.1)** provides reasonably good data in the MS (also showing physically consistent inter-annual variability), while strong negative deviations from the other instruments are found in the USLM. This issue is known and has been related to an imperfect sideband correction of the 488.9 GHz water vapor band.

MIPAS(1) (V3o_H2O_13) and also **MIPAS(2) (V5r_H2O_220)** compare very well to the *MIM* with deviations from the *MIM* mostly within ±5% throughout the atmosphere. An exception is the tropical UTLS (100-300 hPa), where deviations for MIPAS(1) and MIPAS(2) increase to -25% and -10%, respectively. The seasonal cycle and inter-annual variability in the tropical tropopause region exhibit a too low amplitude, which can be explained by a state-dependent averaging kernel. The two data versions agree with each other mostly within a few percent. Exceptions are the UTLS (100-300 hPa), and the tropical LS and US, where MIPAS(1) is about 10% lower than MIPAS(2).

Aura-MLS (V3.3) shows very good to excellent agreement with the *MIM* throughout most of the atmosphere (with deviations from the *MIM* between +2.5 and +5%). Exceptions are found in the LM, where the deviations increase to +10%. Good spatial and temporal coverage (also long-term) allow generally a robust assessment of the Aura-MLS deviations from the *MIM*, which makes the data exceptionally useful for data merging. However, these measurements tend to be low at high latitudes in the lowermost stratosphere (around 200 hPa).

ACE-FTS (V2.2) performs exceptionally well compared to the *MIM* in the tropical and extra-tropical MS and US, the extra-tropical LS, and to a somewhat lesser extent tropical LS and the LM, despite its disadvantage of being an occultation sounder with small temporal and spatial sampling. The deviations from the *MIM* are mostly consistent with validation results using coincident measurements. In the UTLS between 100 and 300 hPa, the deviations from the *MIM* increase to +10% in the extra-tropics and +35% in the tropics, respectively, some of which is likely attributable to limited sampling.

SCIAMACHY (V3.0) water vapor (a relatively new retrieval product) provides promising results, however, suffers from a relatively coarse vertical resolution in the UTLS, which leads to smearing of the strong gradients found across the tropopause when interpolating the data onto the SPARC Data Initiative pressure grid. In this region, the smearing affects mainly the water vapor mean values of SCIAMACHY, while comparison to the other instruments indicate that it does not compromise evaluations of inter-annual variability or amplitudes in water vapor seasonal cycles.

More generally, our evaluations have shown that most sensors exhibit very good agreement in the magnitude and structure in inter-annual variability, after the instruments' biases are removed. The instruments therefore fulfill one of the most important and necessary requirements to be of use for climate variability studies and merging activities, although more investigations will be needed in order to confirm the long-term stability of the datasets, an issue raised in a study by *Fueglistaler et al.* [2013] using a subset of the datasets discussed here. When different datasets are merged for trend studies, the instruments' strength and weaknesses as detailed in *Sections 5.1–5.3* should be taken into account, especially with respect to their performance in different atmospheric regions. An instrument performing badly in the tropical lower stratosphere may show excellent behavior in the mesosphere, or vice-versa. The results also indicate that the combined water vapor datasets show great potential for improving past model-measurement comparisons after accounting for the derived uncertainties. However, careful choices have to be made when choosing instruments to be included in a metric depending on the region of the atmosphere. A region where the requirement of well-defined biases is not met by

most instruments for example is the UTLS (100-300 hPa), emphasizing the need for limb-sounders with higher quality (or consistency) and vertical resolution measurements that reach into this region.

Acknowledgments. The climatology comparisons presented in this study resulted from an activity within the World Climate Research Programme's (WCRP's) Stratospheric Processes and their Role in Climate project – the SPARC Data Initiative. We thank in particular the instrument teams and the various space agencies (CSA, ESA, NASA, and other national agencies) for providing their data and manpower. The SPARC Data Initiative thanks the International Space Science Institute in Bern (ISSI) who supported the activity through their ISSI International Team activity program, and the WCRP and the Toronto SPARC office for generous travel funds. MIH thanks the CSA and ESA for supporting her work for the SPARC Data Initiative. The work from ST was funded from the WGL project TransBrom and the EU project SHIVA (FP7-ENV-2007-1-226224). Work at the Jet Propulsion Laboratory, California Institute of Technology, was performed under contract with NASA. The Atmospheric Chemistry Experiment (ACE), also known as SCISAT, is a Canadian-led mission mainly supported by the CSA. Development of the ACE-FTS climatologies was supported by grants from the Canadian Foundation for Climate and Atmospheric Sciences and the CSA. Work on the SCIAMACHY water vapor at the University of Bremen was supported by the DFG Research Unit FOR 1095 'Stratospheric Change and its role for Climate Prediction (SHARP)' (Project: GZ WE 3647/3-1). Odin is a Swedish-led satellite project funded jointly by Sweden (Swedish National SpaceBoard), Canada (CSA), France (Centre National d'tudes Spatiales), and Finland (Tekes), with support by the 3rd party mission programme of ESA.

References

Bates, D. R., and M. Nicolet (1950), The photochemistry of atmospheric water vapour, *J. Geophys. Res.*, *55*(3), 301327, doi:10.1029/ JZ055i003p00301.

Carleer, M. R., C. D. Boone, K. A. Walker, et al. (2008), Validation of water vapour profiles from the Atmospheric Chemistry Experiment (ACE), *Atmos. Chem. Phys. Discuss.*, *8*, 44994559, http://www.atmos-chem-phys-discuss.net/8/4499/2008/.

Forster, P. M. d. F., and K. P. Shine (2002), Assessing the climate impact of trends in stratospheric water vapor, *Geophys. Res. Lett.*, *29*(6), 1086, 10.1029/2001GL013909.

Fueglistaler, S., and P. H. Haynes (2005), Control of interannual and longer-term variability of stratospheric water vapor, *J. Geophys. Res.*, *110*, D24108, doi:10.1029/2005JD006019.

Fueglistaler, S., A. E. Dessler, T. J. Dunkerton, I. Folkins, Q. Fu, and P. W. Mote (2009), The tropical tropopause layer, *Rev. Geophys.*, *47*, RG1004, doi:10.1029/2008RG000267.

Fueglistaler, S. (2012), Stepwise changes in stratospheric water vapor?, *J. Geophys. Res.*, *117*, D13302, doi:10.1029/2012JD017582.

Fueglistaler, S., Y.S. Liu, T.J. Flannaghan, P.H. Haynes, D.P. Dee, W.J. Read, E.E. Remsberg, L.W. Thomason, D.F. Hurst, J.R. Lanzante, P.F. Bernath (2013), The relation between atmospheric humidity and temperature trends for stratospheric water, *J. Geophys. Res.*, *118*, 1052-1074, doi:10.1002/jgrd.50157.

Fujiwara, M., H. Voemel, F. Hasebe, et al. (2010), Seasonal to decadal variations of water vapour in the tropical lower stratosphere observed with balloon-borne cryogenic frost point hygrometers, *J. Geophys. Res.*, *115*, D18304, doi:10.1029/2010JD014179.

Funke, B. and T. von Clarmann (2012), How to average logarithmic retrievals? *Atmos. Meas. Tech.*, *5*, 831-841.

Gettelman, A., et al. (2010), Multi-model assessment in the upper troposphere and lower stratosphere: tropics and trends, *J. Geophys. Res.*, *115*, D00M08, doi:10.1029/2009JD013638.

Gettelman, A., P. Hoor, L. L. Pan, W. J. Randel, M. I. Hegglin, and T. Birner (2011), The extra-tropical upper troposphere and lower stratosphere, Rev. Geophys., 49, RG3003, doi:10.1029/2011RG000355.

Grooss, J.-U. and J. M. Russell III (2005), Technical note: A stratospheric climatology for O_3, H_2O, CH_4, NO_x, HCl and HF derivd from HALOE measurements, *Atmos. Chem. Phys.*, Vol. 5, 2797-2807.

Gunson, M. R., M. M. Abbas, M. C. Abrams, M. Allen, L. R. Brown, T. L. Brown, A. Y. Chang, A. Goldman, F. W. Irion, L. L. Lowes, E. Mahieu, G. L. Manney, H. A. Michelsen, M. J. Newchurch, C. P. Rinsland, R. J. Salawitch, G. P. Stiller, G. C. Toon, Y. L. Yung, and R. Zander (1996), The Atmospheric Trace Molecule Spectroscopy (ATMOS) experiment: Deployment on the ATLAS Space Shuttle missions, *Geophys. Res. Lett.*, *23*, 2333-2336.

Hegglin, M. I., C. D. Boone, G. L. Manney, T. G. Shepherd, K. A. Walker, P. F. Bernath, W. H. Daffer, P. Hoor, and C. Schiller (2008), Validation of ACE-FTS satellite data in the upper troposphere/lower stratosphere (UTLS) using non-coincident measurements, *Atmos. Chem. Phys.*, *8*, 1483-1499.

Hegglin, M. I., et al. (2010), Multi-model assessment of the upper troposphere and lower stratosphere: extra-tropics, *J. Geophys. Res.*, *115*, D00M09, doi:10.1029/2010JD013884.

Hoor, P., H. Wernli, M. I. Hegglin, and H. Bönisch (2010), Transport timescales and tracer properties in the extra-tropical UTLS, *Atmos. Chem. Phys.*, *10*, 7929-7944, doi:10.5194/acp-10-7929-2010.

Hurst, D. F., S. J. Oltmans, H. Vmel, K. H. Rosenlof, S. M. Davis, E. A. Ray, E. G. Hall, and A. F. Jordan (2011), Stratospheric water vapour trends over Boulder, Colorado: Analysis of the 30 year Boulder record, *J. Geophys. Res.*, *116*, D02306, doi:10.1029/2010JD015065.

IPCC (2007), Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B.M.Tignor and H.L. Miller (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 996 pp.

Jiang, J.H., H. Su, S. Pawson, H.C. Liu, W. Read, J.W. Waters, M. Santee, D.L. Wu, M. Schwartz, N. Livesey, A. Lambert, R. Fuller, and J.N. Lee (2010), Five-year (2004-2009) Observations of Upper Tropospheric Water Vapor and Cloud Ice from MLS and Comparisons with GEOS-5 analyses, *J. Geophys. Res., 115*, D15103, doi:10.1029/2009JD013256.

Jiang, J.H., H. Su, C. Zhai, V.S. Perun, A. Del Genio, L.S. Nazarenko, L.J. Donner, L. Horowitz, C. Seman, J. Cole, A. Gettelman, M. Ringer, L. Rotstayn, S. Jeffrey, T. Wu, F. Brient, J-L. Dufresne, H. Kawai, T. Koshiro, M. Watanabe, T.S. Lcuyer, E.M. Volodin, T. Iversen, H. Drange, M.S. Mesquita, W.G. Read, J.W. Waters, B. Tian, J. Teixeira, and G.L. Stephens (2012), Evaluation of Cloud and Water Vapor Simulations in CMIP5 Climate Models Using NASA A-Train Satellite Observations, *J. Geophys. Res., 117*, D1410, 24 PP, doi:10.1029/2011JD017237.

Jones, A., K. A. Walker, J. J. Jin, J. R. Taylor, C. D. Boone, P. F. Bernath, S. Brohede, G. L. Manney, S. McLeod, R. Hughes, and W. H. Daffer (2012), Technical Note: A trace gas climatology derived from the Atmospheric Chemistry Experiment Fourier Transform Spectrometer dataset, *Atmos. Chem. Phys.*, *12*, 5207-5220, doi:10.5194/acp-12-5207-2012.

Kelly, K. K., et al. (1989), Dehydration in the lower antarctic stratosphere during late winter and early spring 1987, *J. Geophys. Res.*, *94*, 11317–11357.

Lambert, A., et al. (2007), Validation of the Aura Microwave Limb Sounder middle atmosphere water vapour and nitrous oxide measurements, *J. Geophys. Res.*, *112*, D24S36, doi:10.1029/2007JD008724.

Livesey, N. J., et al. (2011), EOS MLS Version 3.3 Level 2 data quality and description document, JPL D-33509, Jet Propulsion Laboratory, California Institute of Technology.

Lucke, R. L., et al. (1999), The Polar Ozone and Aerosol Measurement (POAM III) Instrument and Early Validation Results, *J. Geophys. Res.*, *104*, 18,785-18,799.

Lumpe, J., et al. (2006), Validation of Polar Ozone and Aerosol Measurement (POAM) III version 4 stratospheric water vapour, *J. Geophys. Res.*, *111*, D11301, doi:10.1029/2005JD006763.

Milz, M. et al. (2009), Validation of water vapour profiles (version 13) retrieved by the IMK/IAA scientific retrieval processor based on full resolution spectra measured by MIPAS on board Envisat, *Atmos. Meas. Techn.*, *2*, 379399.

Milz, M., T. von Clarmann, H. Fischer, N. Glatthor, U. Grabowski, M. Hpfner, S. Kellmann, M. Kiefer, A. Linden, G. Mengistu Tsidu, T. Steck, G. P. Stiller, B. Funke, M. Lpez-Puertas, and M. E. Koukouli (2005), Water vapour distributions measured with the Michelson Interferometer for Passive Atmospheric Sounding on board Envisat (MIPAS/Envisat). *J. Geophys. Res.*, *110*, doi:10.1029/2005JD005973.

Mote, P.W., K.H. Rosenlof, M.E. McIntyre, et al. (1996), An atmospheric tape recorder: The imprint of tropical tropopause temperatures on stratospheric water vapour, *J. Geophys.Res.*, *101*, 3989-4006.

Nedoluha, G. E., et al. (2007), A comparison of middle atmospheric water vapour as measured by WVMS, E_S-MLS, and HALOE, *J. Geophys. Res.*, *112*, D24S39, doi:10.1029/2007JD008757.

Niwano, M., K. Yamazaki, and M. Shiotani (2003), Seasonal and QBO variations of ascent rate in the tropical lower stratosphere as inferred from UARS HALOE trace gas data, *J. Geophys. Res.*, *108*(D24), 4794, doi:10.1029/2003JD003871.

Pumphrey, H. C. (1999), Validation of a new prototype water vapour retrieval fo the UARS Microwave Limb Sounder, *J. Geophys. Res.*, *104*(D8), 9399-9412.

Randel, W.J., F. Wu, S. Oltmans, K. Rosenlof and G. Nedoluha (2004), Inter-annual changes of stratospheric water vapor and correlations with tropical tropopause temperatures, J. *Atmos. Sci.*, *61*, 2133-2148.

Read, W. G., et al. (2007), Aura Microwave Limb Sounder upper tropospheric and lower stratospheric H2O and relative humidity with respect to ice validation, *J. Geophys. Res.*, *112*, D24S35, doi:10.1029/2007JD008752.

Remsberg, E. E., M. Natarajan, G. S. Lingenfelser, R. E. Thompson, B. T. Marshall, and L. L. Gordley (2009), On the quality of the Nimbus 7 LIMS version 6 water vapour profiles and distributions, *Atmos. Chem. Phys.*, *9*, 9155-9167, doi:10.5194/acp-9-9155-2009.

Roche, A. E., J. B. Kumer, J. L. Mergenthaler, G. A. Ely, W. G. Uplinger, J. F. Potter, T. C. James, and L. W. Sterritt (1993), The Cryogenic Limb Array Etalon Spectrometer (CLAES) on UARS: Experiment Description and Performance, *J. Geophys. Res.*, *98*(D6), 10,76310,775, doi:10.1029/93JD00800.

Rousseeuw, P. J. and Croux, C. (1993), Alternatives to the median absolute deviation, *J. Amer. Statist. Assoc.*, *88*, 1273-1283.

Rozanov, A., Weigel, K., Bovensmann, H., Dhomse, S., Eichmann, K.-U., Kivi, R., Rozanov, V., Vmel, H., Weber, M., and Burrows, J. P. (2011), Retrieval of water vapor vertical distributions in the upper troposphere and the lower stratosphere from SCIAMACHY limb measurements, *Atmos. Meas. Tech.*, *4*, 933-954, doi:10.5194/amt-4-933-2011.

Seidel, D. J., N. P. Gillett, J. R. Lanzante, K. P. Shine, and P. W. Thorne (2011), Stratospheric temperature trends: Our evolving understanding, *Wiley Interdiscip. Rev. Clim. Change*, *2*(4), 592616, doi:10.1002/wcc.125.

Seinfeld, J. H. and S. N. Pandis (2006), Atmospheric Chemistry and Physics - From Air Pollution to Climate Change (2nd Edition), John Wiley & Sons.

Solomon, S. (1999), Stratospheric ozone depletion: A review of concepts and history, *Rev. Geophys.*, *37*, 3, doi:10.1029/1999RG900008.

SPARC WAVAS (2000), Upper Tropospheric and Stratospheric Water Vapour, D. Kley, J. M. Russell III, and C. Phillips (Eds.), WMO/TD-No. 1043, *SPARC Report No. 2.*

SPARC CCMVal (2010), SPARC Report on the Evaluation of Chemistry-Climate Models, V. Eyring, T. G. Shepherd, D. W. Waugh (Eds.), SPARC Report No. 5, WCRP-132, WMO/TD-No. 1526.

Stiller G.P., M. Kiefer, E. Eckert, T. von Clarmann, S. Kellmann, M. Garca-Comas, B. Funke, T. Leblanc, E. Fetzer, L. Froidevaux, M. Gomez, E. Hall, D. Hurst, A. Jordan, N. Kämpfer, A. Lambert, I. S. McDermid, T.McGee, L. Miloshevich, G. Nedoluha, W. Read, M. Schneider, M. Schwartz, C. Straub, G. Toon, L. W. Twigg, K. Walker, D.N.

Whiteman (2012), Validation of MIPAS IMK/IAA temperature, water vapour, and ozone profiles with MOHAVE-2009 campaign measurements, *Atmos. Meas. Techn.*, *5*, 289-320, doi:10.5194/amt-5-289-2012.

Taha, G., L. W. Thomason, and S. P. Burton (2004), Comparison of Stratospheric Aerosol and Gas Experiment (SAGE) II version 6.2 water vapour with balloon-borne and space-based instruments, *J. Geophys. Res.*, *109*, D18313, doi:10.1029/2004JD004859.

Taylor, F. W., C. D. Rodgers, J. G. Whitney, S. T. Werrett, J. J. Barnett, G. D. Peskett, P. Venters, J. Ballard, C. W. P. Palmers, R. J. Knight, P. Morris, T. Nightingale, and A. Dudhia (1993), Remote Sensing of Atmospheric Structure and Composition by Pressure Modulator Radiometry From Space: The ISAMS Experiment on UARS, *J. Geophys. Res.*, *98*(D6), 10,79910,814, doi:10.1029/92JD03029.

Taylor, K. E. (2001), Summarizing multiple aspects of model performance in a single diagram, *J. Geophys. Res.*, *106*, 71837192, doi:10.1029/2000JD900719

Thomason, L. W., S. P. Burton, N. Iyer, J. M. Zawodny, and J. Anderson 2004, A revised water vapour product for the Stratospheric Aerosol and Gas Experiment (SAGE) II version 6.2 dataset, *J. Geophys. Res.*, *109*, D06312, doi:10.1029/2003JD004465.

Thomason, L. W., J. R. Moore, M. C. Pitts, J. M. Zawodny, and C. W. Chiou (2010), An evaluation of the SAGE III version 4 aerosol extinction coefficient and water vapour data products, *Atmos. Chem. Phys.*, *10*, 2159-2173.

Toohey, M. and von Clarmann, T. (2013), Climatologies from satellite measurements: the impact of orbital sampling on the standard error of the mean, *Atmos. Meas. Tech.*, *6*, 937-948, doi:10.5194/amt-6-937-2013.

Urban, J., N. Lauti, D. P. Murtagh, P. Eriksson, Y. Kasai, S. Lossow, E. Dupuy, J. De La No, U. Frisk, M. Olberg, E. Le Flochmon, P. Ricaud (2007), Global observations of middle atmospheric water vapour by the Odin satellite: An overview, *Planetary and Space Science*, *55*(9), pp. 1093-1102.

Urban, J. (2008), Tropical ascent of lower stratospheric air analysed using measurements of the Odin Sub-Millimetre Radiometer. Proc. Reunion Island Int. Symp. Tropical Stratosphere Upper Troposphere, 5-9 November 2007, St. Gilles, Reunion Island, France, Editor: H. Bencherif, Publisher: Universite de la Reunion, pp. 29-34.

Urban, J., D.P. Murtagh, G. Stiller and K.A. Walker (2012), Evolution and Variability of Water Vapour in the Tropical Tropopause and Lower Stratosphere Region Derived from Satellite Measurements, *Proc. ATMOS 2012; Advances in Atmospheric Science and Applications*, ESA SP-708 (CD-ROM), ESA Communications, European Space Agency, Noordwijk, The Netherlands.

Voemel, H., et al. (2007), Validation of Aura Microwave Limb Sounder water vapour by balloon-borne Cryogenic Frost Point Hygrometer measurements, *J. Geophys. Res.*, *112*, D24S37, doi:10.1029/2007JD008698.

von Clarmann, T., M. Höpfner, S. Kellmann, A. Linden, S. Chauhan, B. Funke, U. Grabowski, N. Glatthor, M. Kiefer, T. Schieferdecker, G. P. Stiller, and S. Versick (2009), Retrieval of temperature, H_2O, O_3, HNO_3, CH_4, N_2O, $ClONO_2$ and ClO from MIPAS reduced resolution nominal mode limb emission measurements, *Atmos. Meas. Techn.*, *2*, 159175.

Weinstock, E. M., et al. (2009), Validation of the Harvard Lyman-a in situ water vapour instrument: Implications for the mechanisms that control stratospheric water vapour, *J. Geophys. Res.*, *114*, D23301, doi:10.1029/2009JD012427.

Table 1. Name, full name, satellite platform, observation geometry and measurement wavelength of instruments participating in the SPARC Data Initiative.

Instrument	Full name	Satellite platform	Observation geometry	Measurement wavelength
LIMS	Limb Infrared Monitor of the Stratosphere	Nimbus 7	Limb emission	Mid-IR
SAGE II, III	Stratospheric Aerosol and Gas Experiment	ERBS, Meteor-3M	Solar occultation	Near-IR
UARS-MLS	UARS-Microwave Limb Sounder	UARS	Limb emission	Microwave/Sub-mm
HALOE	The Halogen Occultation Experiment	UARS	Solar Occultation	Mid-IR
POAM III	Polar Ozone and Aerosol Measurement	SPOT-4	Solar occultation	Near-IR VIS/UV
SMR	Sub-Millimetre Radiometer	Odin	Limb emission	Microwave/Sub-mm
MIPAS	Michelson Interferometer for Passive Atmospheric Sounding	Envisat	Limb emission	Mid-IR
SCIAMACHY	Scanning Imaging Absorption spectroMeter for Atmospheric CHartographY	Envisat	Limb scattering	Near-IR VIS/UV
ACE-FTS	Atmospheric Chemistry Experiment - Fourier Transform Spectrometer	SCISAT-1	Solar occultation	Mid-IR
Aura-MLS	Aura-Microwave Limb Sounder	Aura	Limb emission	Microwave/Sub-mm

Table 2. Water vapor datasets used within the SPARC Data Initiative, including information on data version, temporal coverage, vertical extent, vertical resolution of the native retrieval grid, relevant references and other relevant remarks.

Instrument and data version	**Time period**	**Vertical Range**	**Vertical resolution**	**References**	**Additional comments**
LIMS V6.0	Nov 78-May 79	cloud top-1 hPa	3.7 km	Remsberg et al., 2009	–
SAGE II V6.2	Oct 84-Aug 05	cloud top-50 km	1-2.5 km	Thomason et al., 2004	Use for trend studies
		< 25 km	1 km	Taha et al., 2004	not recommended; Data
		> 30 km	2.5 km		above 3 hPa excluded
UARS-MLS V6	Oct 91-Mar 93	18-50 km	3-4 km	Pumphrey, 1999	H_2O stops early due
		> 50 km	5-7 km		to radiometer failure
HALOE V19	Oct 91-Nov 05	cloud top-90 km	2.5 km	Grooss and Russell, 2005	Data below tropopause
					are excluded
SAGE III V4.0	May 02-Dec 05	cloud top-50 km	1.5 km	Thomason et al., 2010	Only solar products
					are used here
POAM III V4.0	Apr 98-Dec 05	5-45 km	1-2 km	Lumpe et al., 2006	–
				Lucke et al., 1999	
SMR	Jul 01-	16-75 km			
SMR(2) V2.0		16-20 km	3-4 km	Urban, 2008; 2012	544 GHz-band
SMR(1) V2.1		20-75 km	3 km	Urban et al., 2007; 2012	489 GHz-band
MIPAS		cloud top-70 km			Measurement mode switched
MIPAS(1)	Mar 02-Mar 04		4.5-6.5 km	Milz et al., 2005	in 2005 from high spectral
V3o_H2O_13					to high vertical resolution
MIPAS(2)	Jan 05-Apr 12		2.5-6.9 km	Milz et al., 2009	
V5r_H2O_220				von Clarmann et al., 2009	
SCIAMACHY	Sep 02-Apr 12	11-25 km	3-5 km	Rozanov et al., 2011	New data product; note
V3.0					reduced latitudinal coverage
					during winter ($<55°$S/N);
					only every 8th day included
ACE-FTS V2.2	Mar 04-	5-89 km	3-4 km	Carleer et al., 2008	–
				Hegglin et al., 2008	
Aura-MLS V3.3	Aug 04-	316-100 hPa	2-3 km	Read et al., 2007	–
		100-0.2 hPa	3-4 km	Lambert et al., 2007	–
		< 0.1 hPa	6-12 km	Livesey et al., 2011	–

Table 3. Definitions and abbreviations of different atmospheric regions as used in this study.

Note, the full height range corresponds to about 9-65 km.

Region	**Abbreviation**	**Lower boundary**	**Upper boundary**
Upper Troposphere	**UT**	300 hPa	tropopause
Lower Stratosphere	**LS**	tropopause	30 hPa
Middle Stratosphere	**MS**	30 hPa	5 hPa
Upper Stratosphere	**US**	5 hPa	1 hPa
Lower Mesosphere	**LM**	1	0.1 hPa

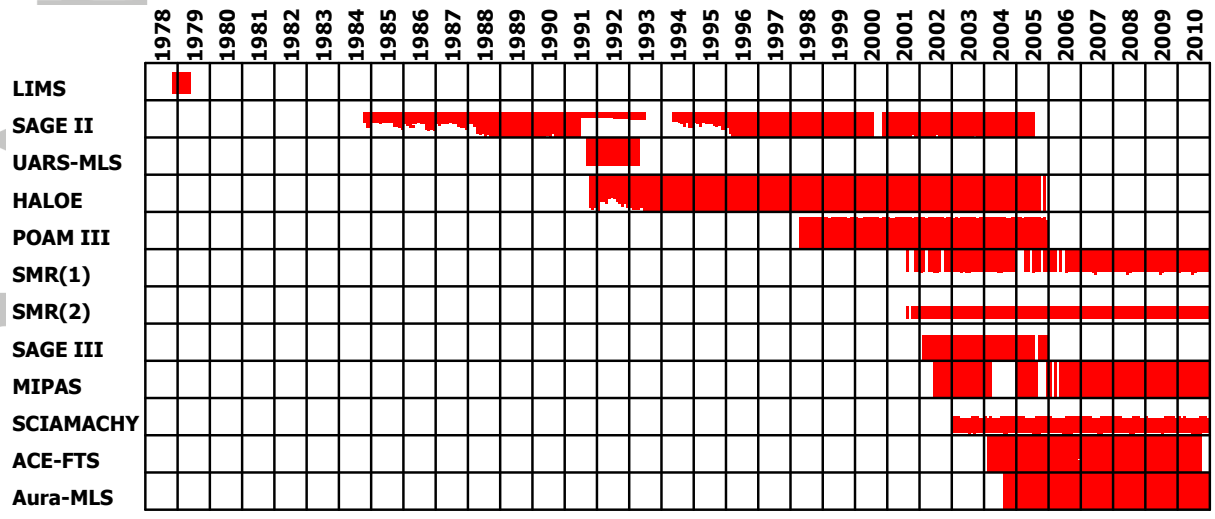

Figure 1. Available water vapor data records between 1978 and 2010 from limb-sounding satellite instruments participating in the SPARC Data Initiative. The red filling of the grid boxes indicates the temporal (January to December) and vertical coverage (300 to 0.1 hPa) of the respective instrument in a given year.

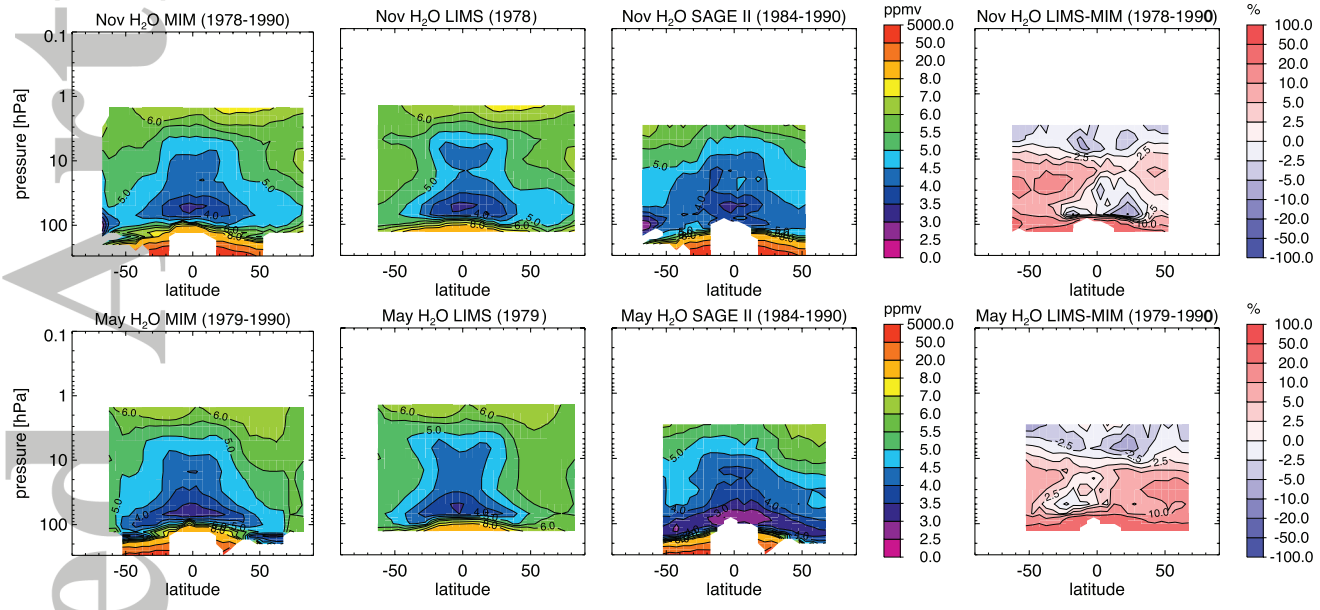

Figure 2. Cross-sections of November and May monthly zonal mean water vapor for the *MIM*, LIMS, and SAGE II (from left to right) for the years indicated in the headings. Right panels show the relative differences between LIMS and the *MIM*. Note that SAGE II differences have the same values, but are of opposite sign.

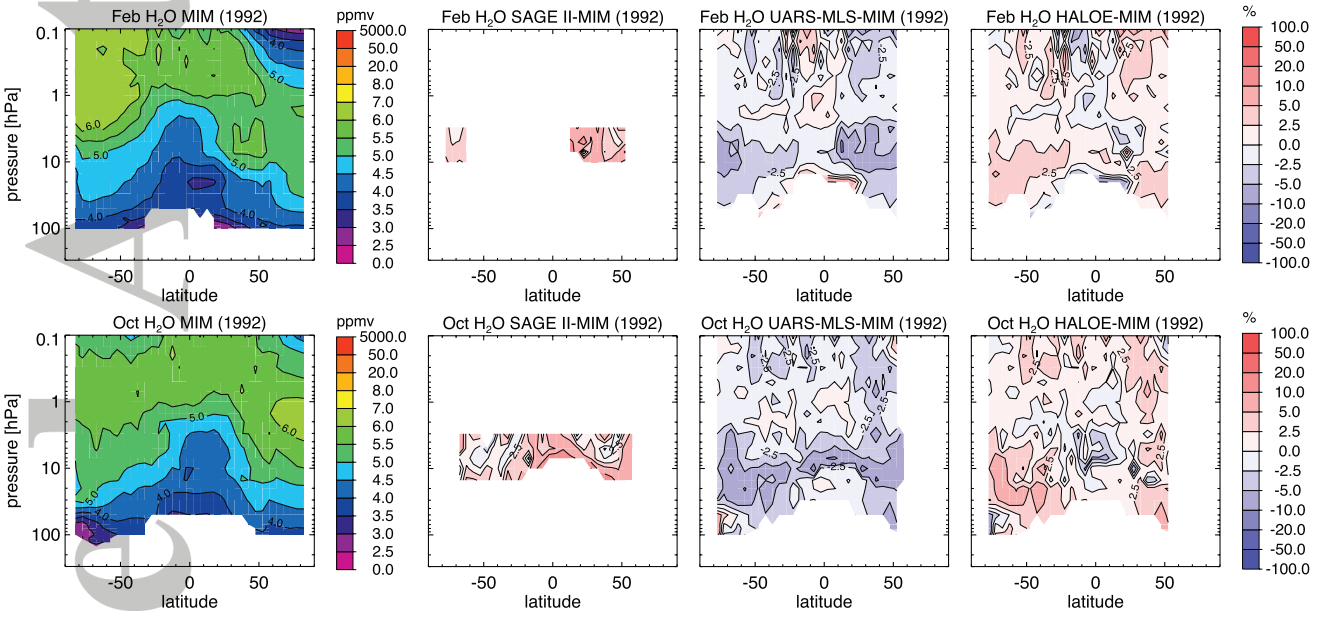

Figure 3. Cross-sections of February and October monthly zonal mean water vapor for the *MIM* in 1992 (left panels). Right panels show the relative differences between SAGE II, UARS-MLS, and HALOE with respect to the *MIM*, respectively.

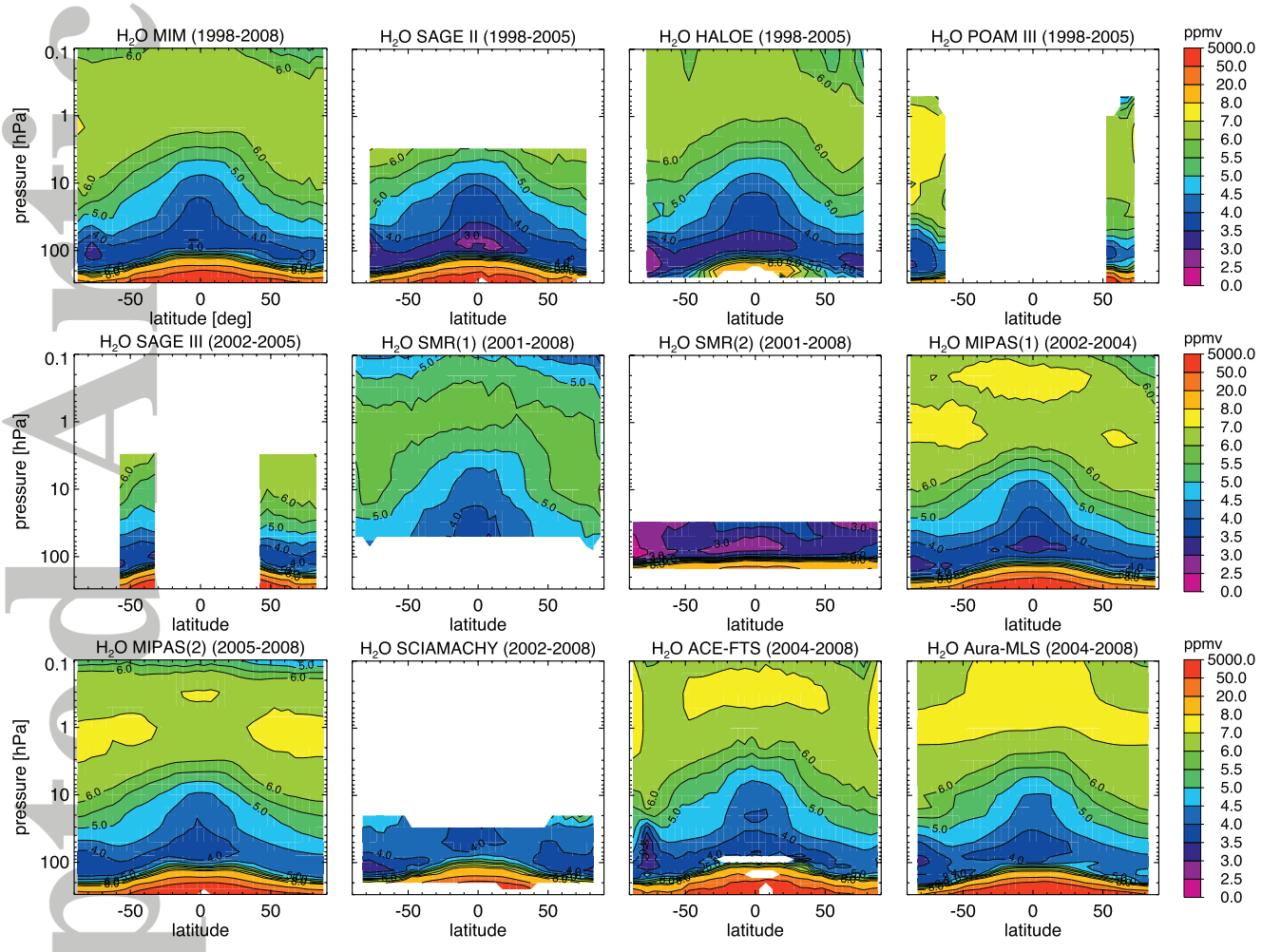

Figure 4. Annual zonal mean cross-sections of water vapor for the *MIM* and the individual instruments averaged over the time period 1998-2008. Note that SMR(2) and MIPAS(1) are not included in the *MIM*.

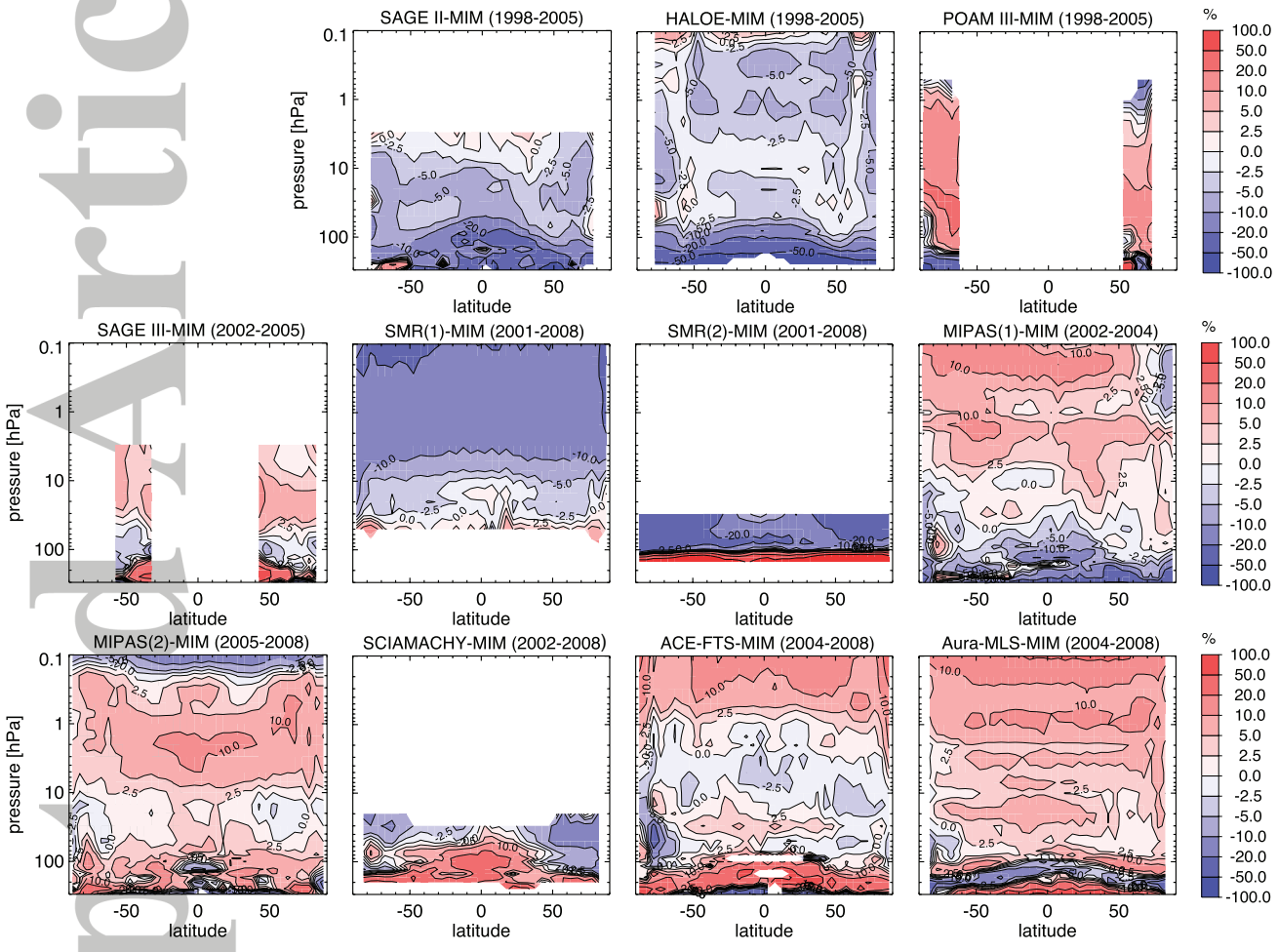

Figure 5. Relative difference cross-sections with respect to the *MIM* for each individual instruments' water vapor distribution shown in Figure 4. Note that SMR(2) and MIPAS(1) are not included in the *MIM*.

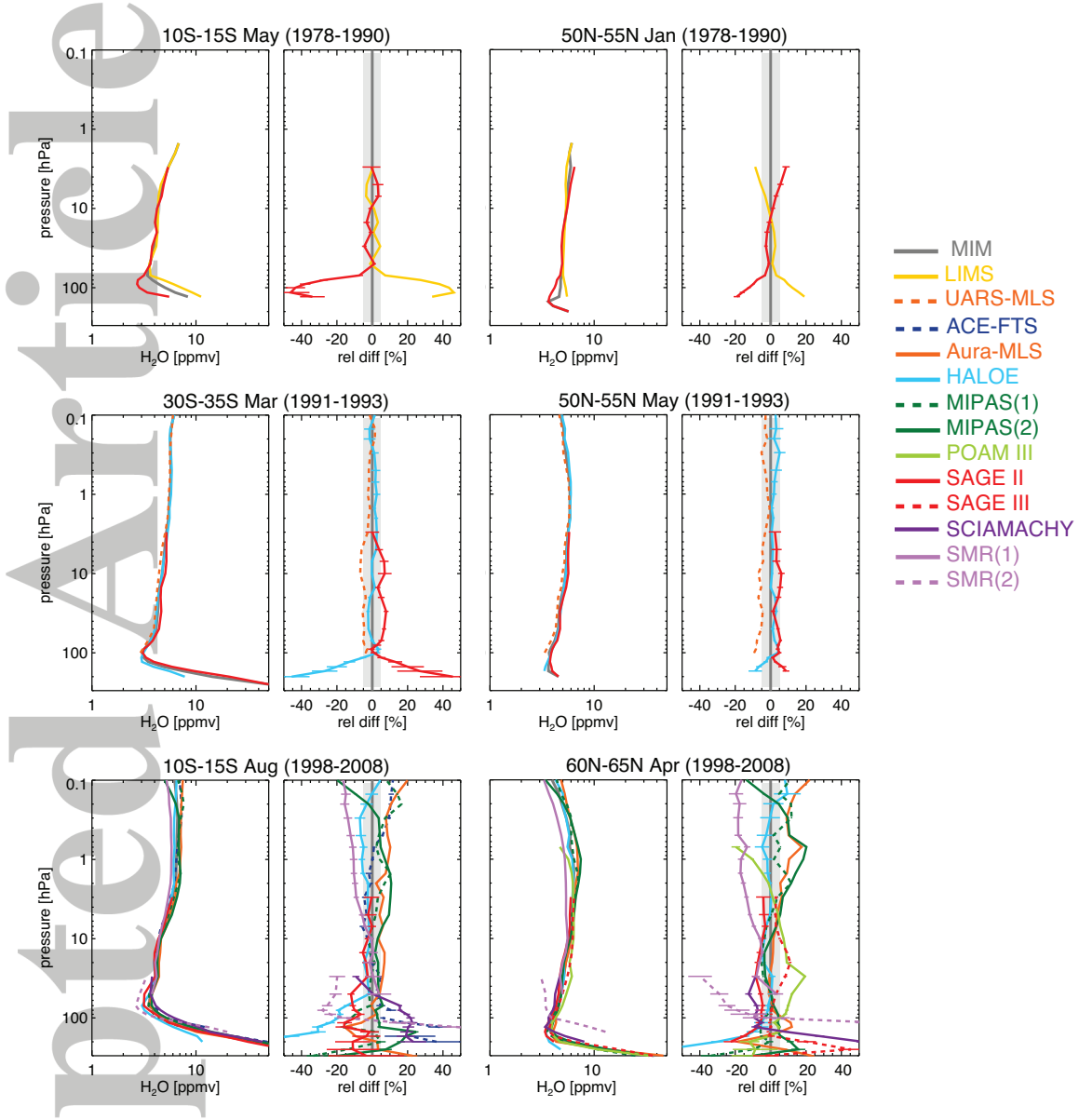

Figure 6. Vertical profiles of monthly zonal mean water vapor for three different periods (1978-1990, 1991-1993, and 1998-2008; top to bottom). Left panels show absolute values, right panels relative differences between the individual instruments and the *MIM*. The grey shaded area indicates where relative differences are smaller than 5%. Bars indicate the uncertainties in the relative differences based on the *SEM* of each instrument.

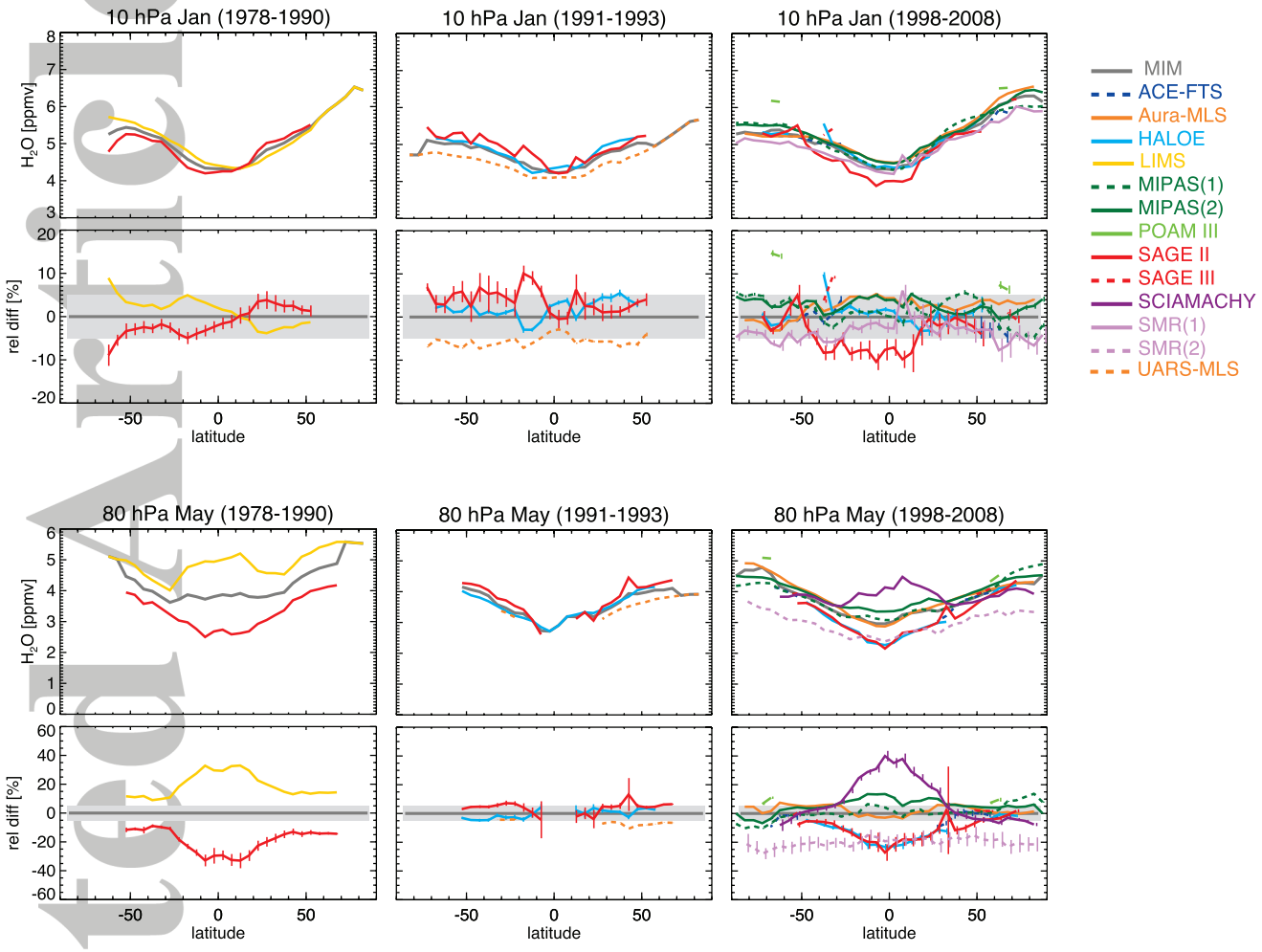

Figure 7. Meridional profiles of monthly mean water vapor for three different time periods (1978-1990, 1991-1993, and 1998-2008; left to right) are shown at 10 (upper row) and 80 (lower row) hPa. Upper panels show absolute values, lower panels relative differences between the individual instruments and the *MIM*, respectively. The grey shading indicates where the relative differences are smaller than 5%. Bars indicate the uncertainties in the relative differences based on the *SEM* of each instrument.

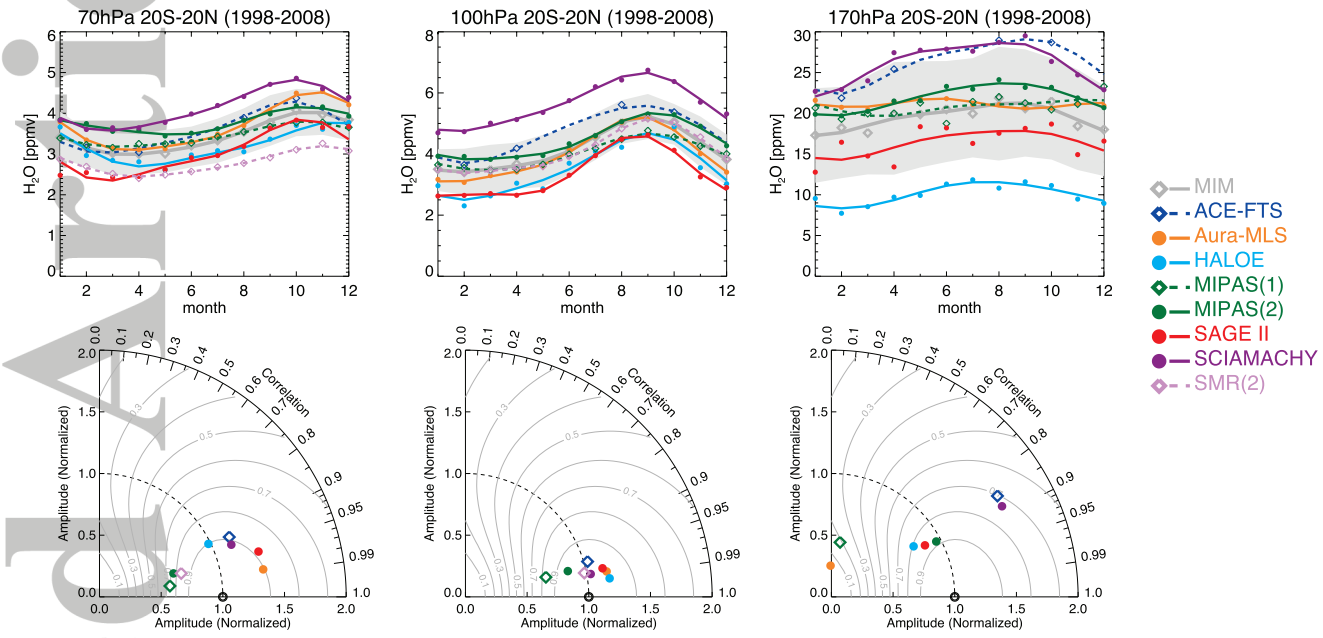

Figure 8. Seasonal cycles of water vapor in the tropics for 1998-2008. Shown are seasonal cycles (top panels) and corresponding Taylor diagrams (lower panels) of monthly zonal mean water vapor averaged over 20°S to 20°N at 70, 100, and 170 hPa (from left to right). The dots indicate actual monthly mean values and the colored lines represent fits including an annual and a semi-annual component to the available monthly data points. The grey line indicates the multi-instrument mean (*MIM*) and the grey shading 1σ.

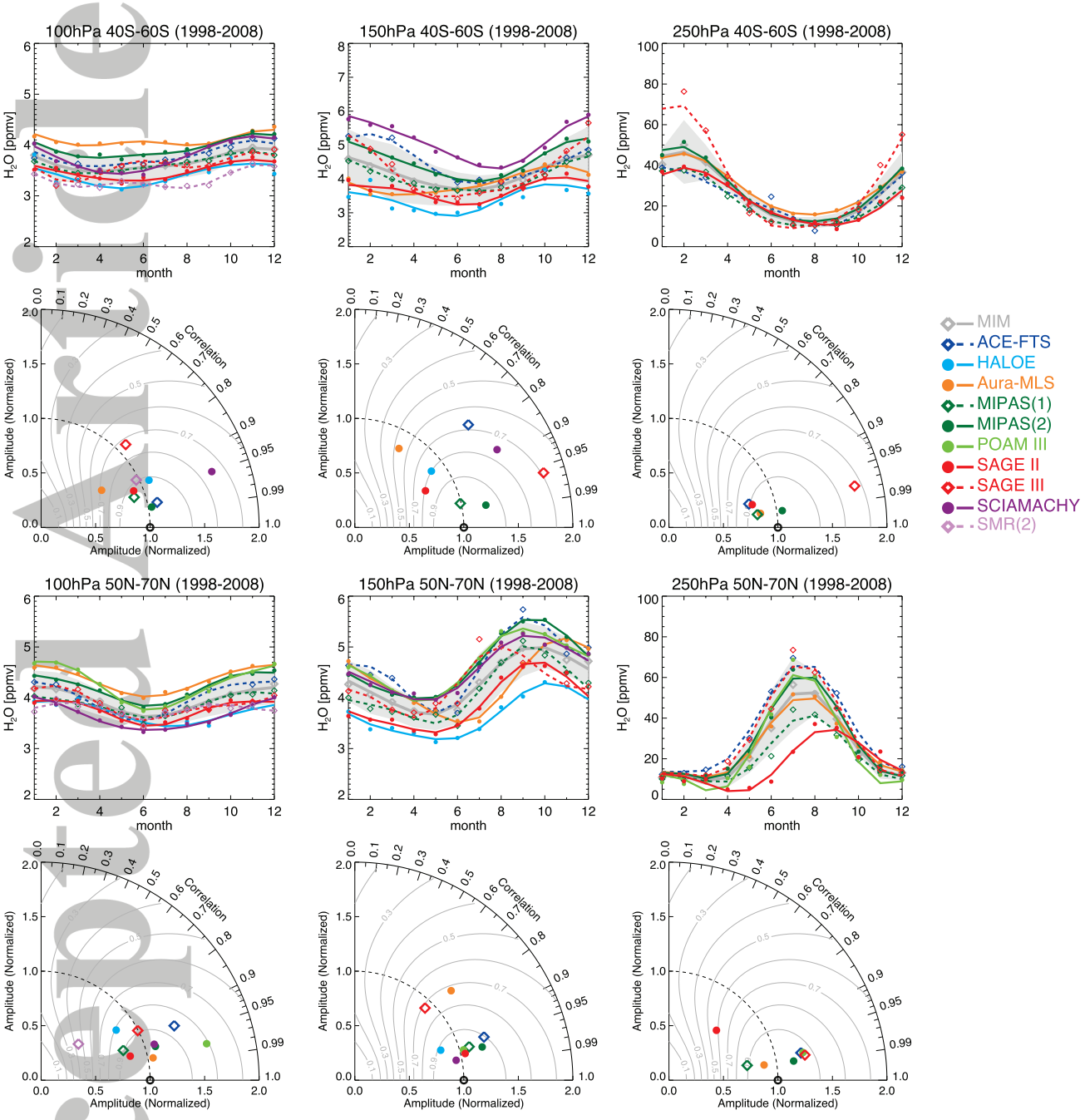

Figure 9. Same as Figure 8, but for 100, 150, and 250 hPa (from left to right) for the extra-tropics between 40°S and 60°S (upper two rows) and between 50°N to 70°N (lower two rows).

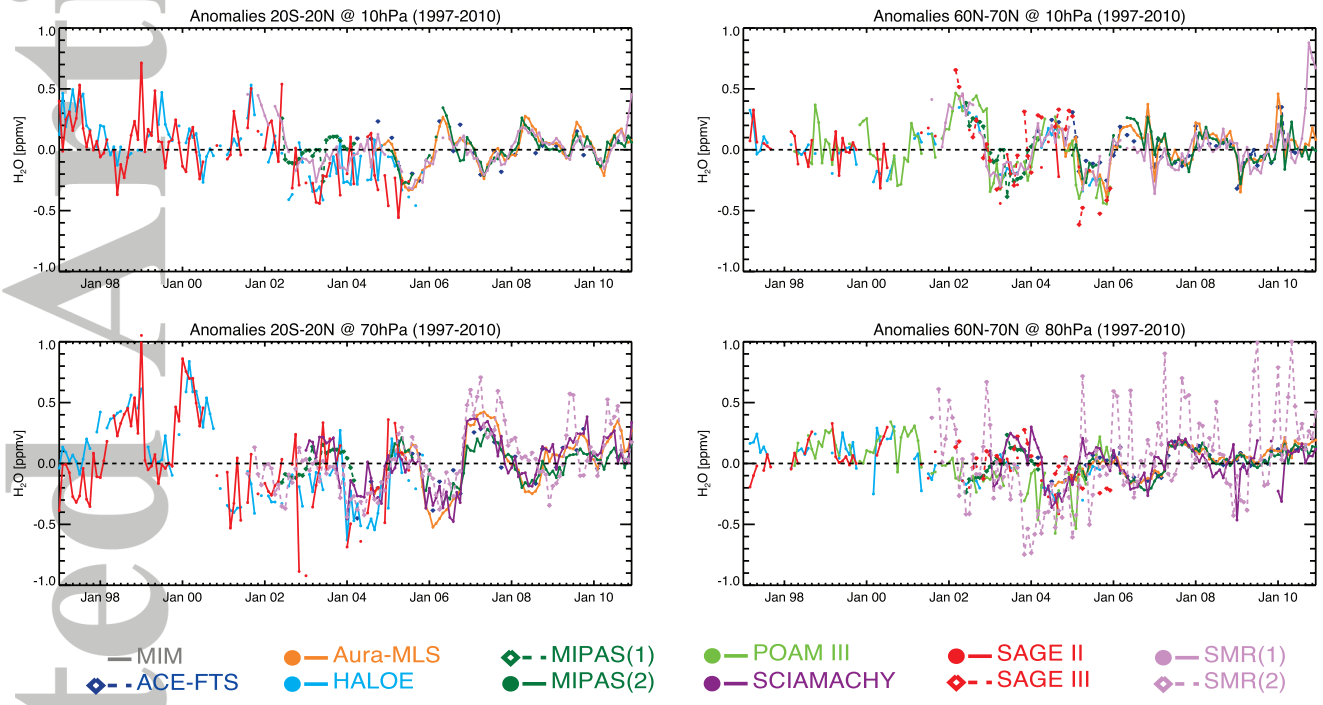

Figure 10. Time series of deseasonalized water vapor anomalies at 10 and 70 hPa in the tropics (20°S-20°N) (left panels) and at 10 and 80 hPa in the extra-tropics (60°N-70°N) (right panels) between 1997 and 2010.

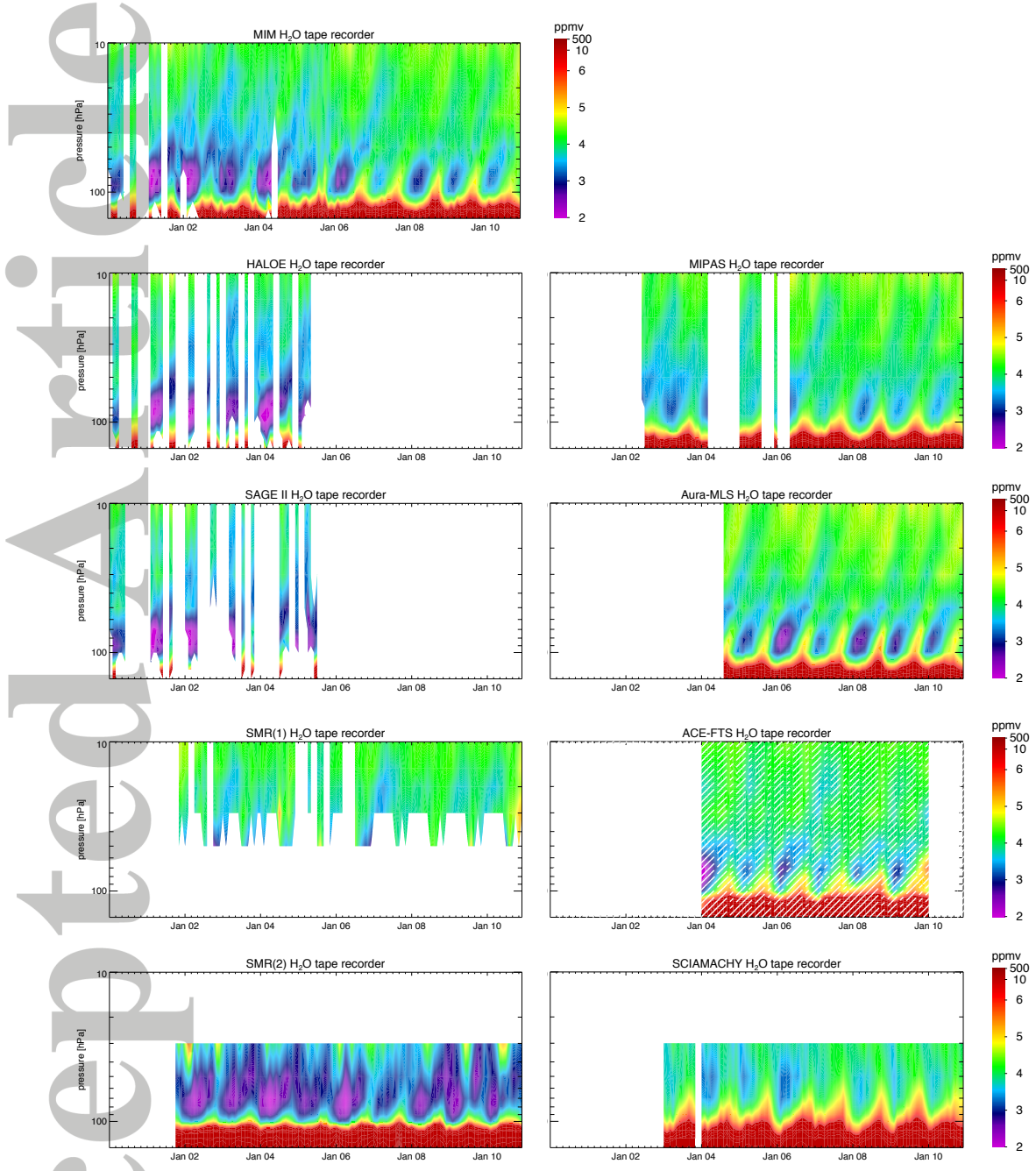

Figure 11. The tape recorder (altitude-time evolution) of water vapor averaged over 15°S-15°N for the time period 2000-2010 is shown for the *MIM* (uppermost left panel) and the different instruments. Note, the very limited ACE-FTS data in the tropics were interpolated in time and altitude with white hatching indicating regions that do not contain data.

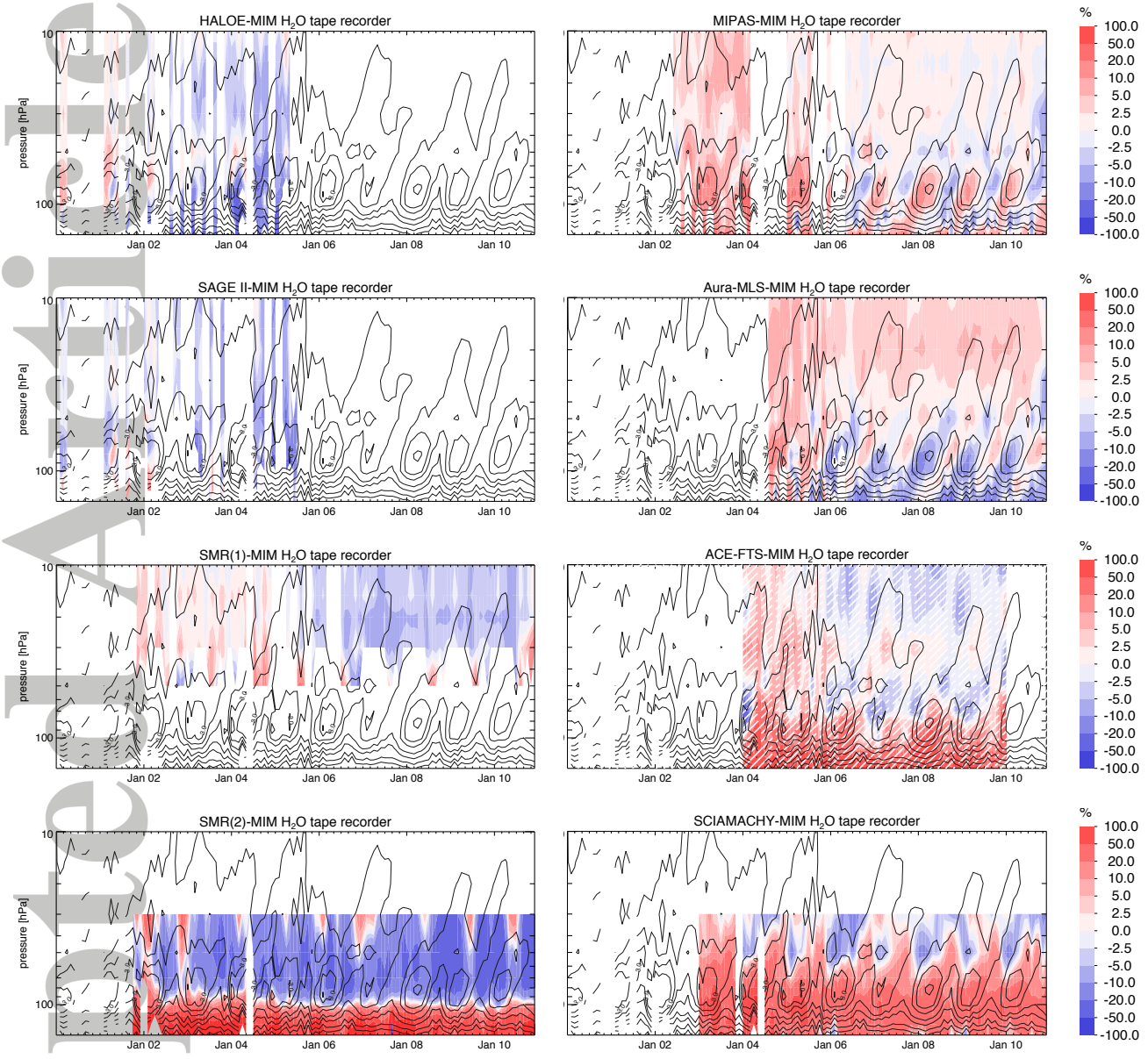

Figure 12. Shown are the differences between each individual instrument's tape recorder and the *MIM* as seen in Figure 11. Contour levels (2.5, 3, 3.5, 4, 5, 6, 8, 10, 50, 100 ppmv, with the 3-ppmv isopleths labeled) reproduce the *MIM*. Note, the very limited ACE-FTS data in the tropics were interpolated in time and altitude with white hatching indicating regions that do not contain data.

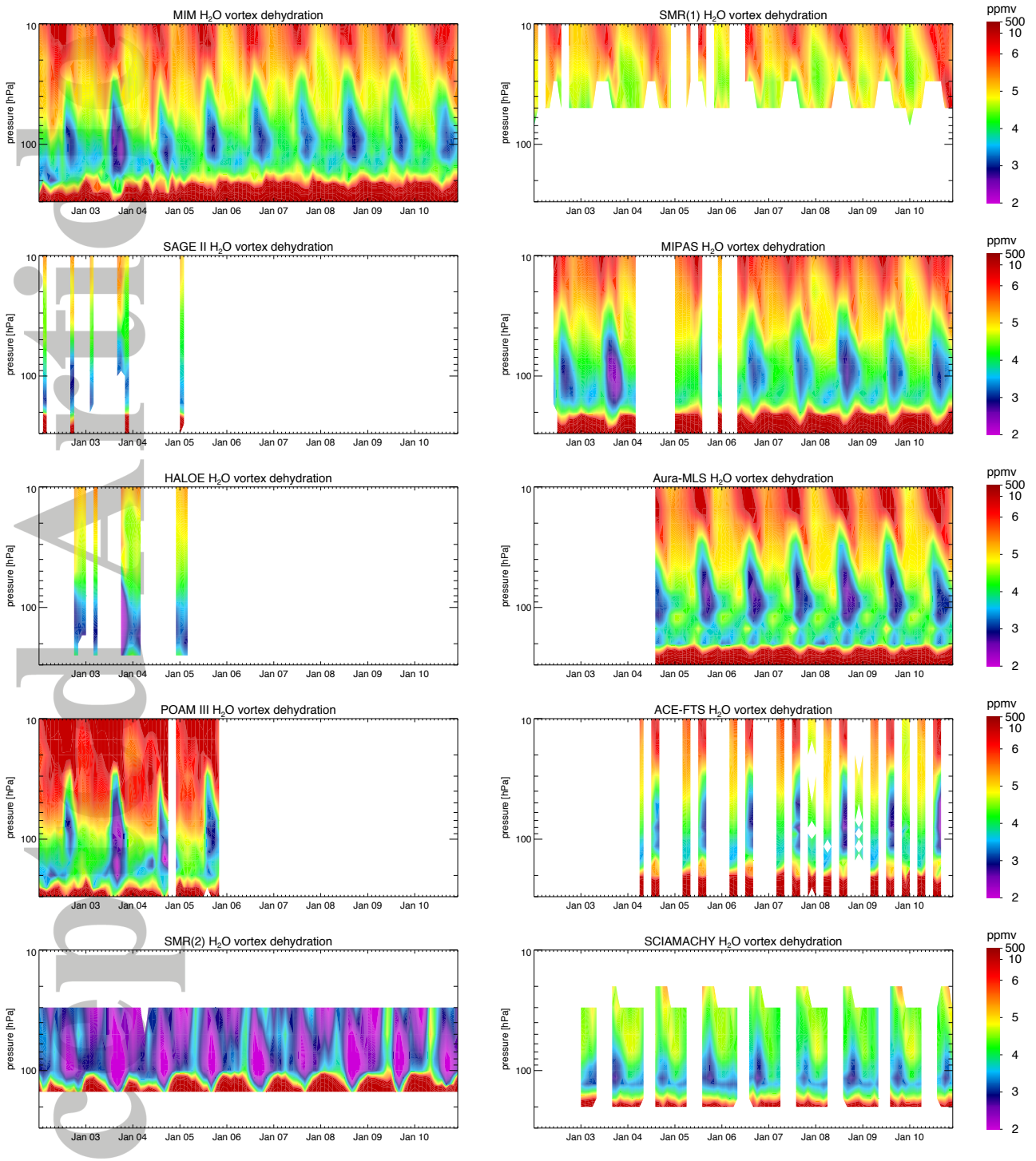

Figure 13. The altitude-time evolution of Antarctic polar vortex descent and dehydration between 2002 and 2010 is shown for the *MIM* (uppermost panel) and the different instruments using water vapor averaged over 60°S to 90°S. Note that SMR(2) is not included in the *MIM*.

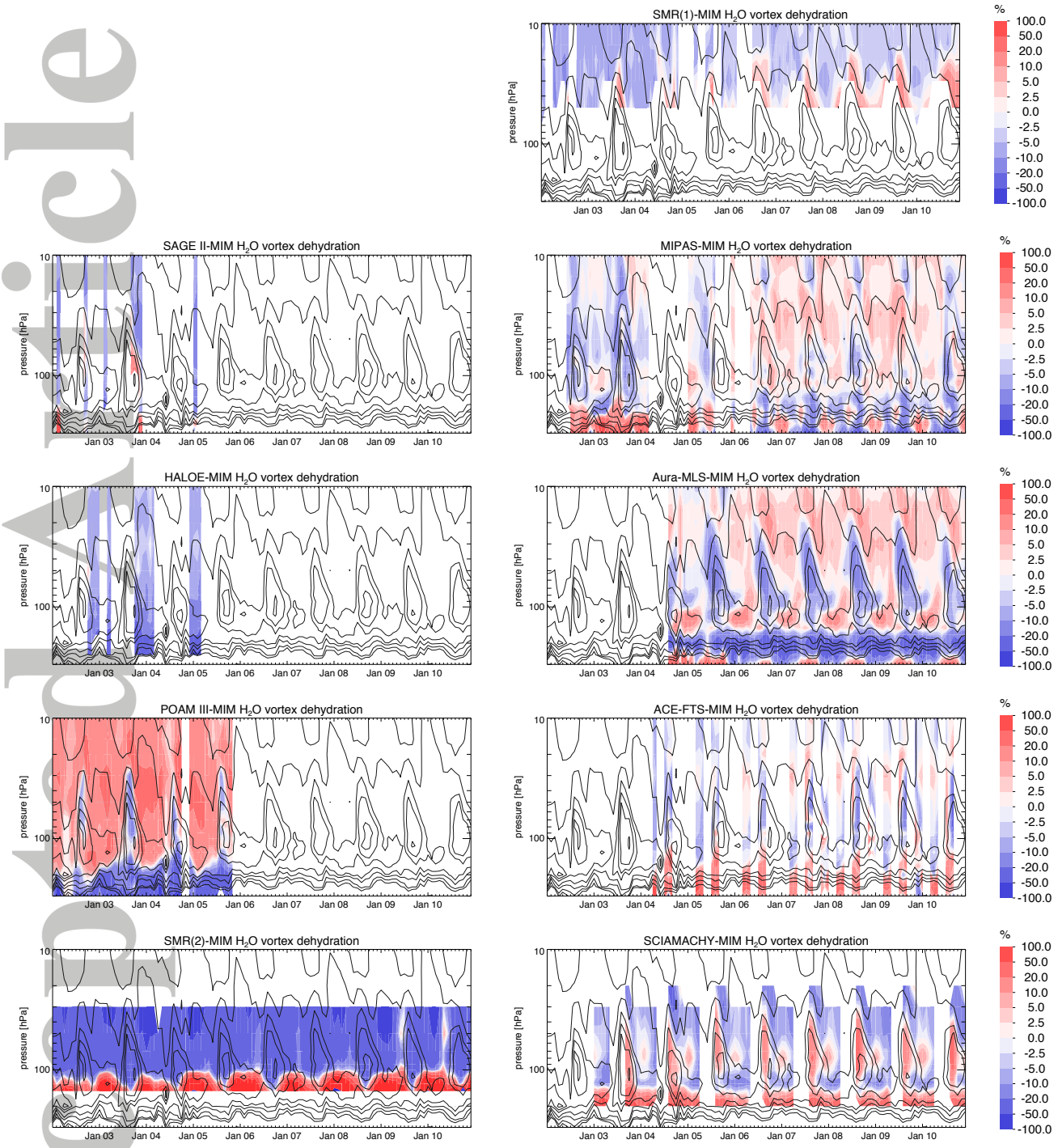

Figure 14. Shown are the relative differences between each individual instrument and the *MIM* as seen in Figure 13. Contour levels (2.5, 3, 3.5, 4, 5, 6, 8, 10, 50, 100 ppmv, with the 3-ppmv isopleths labeled) reproduce the *MIM*.

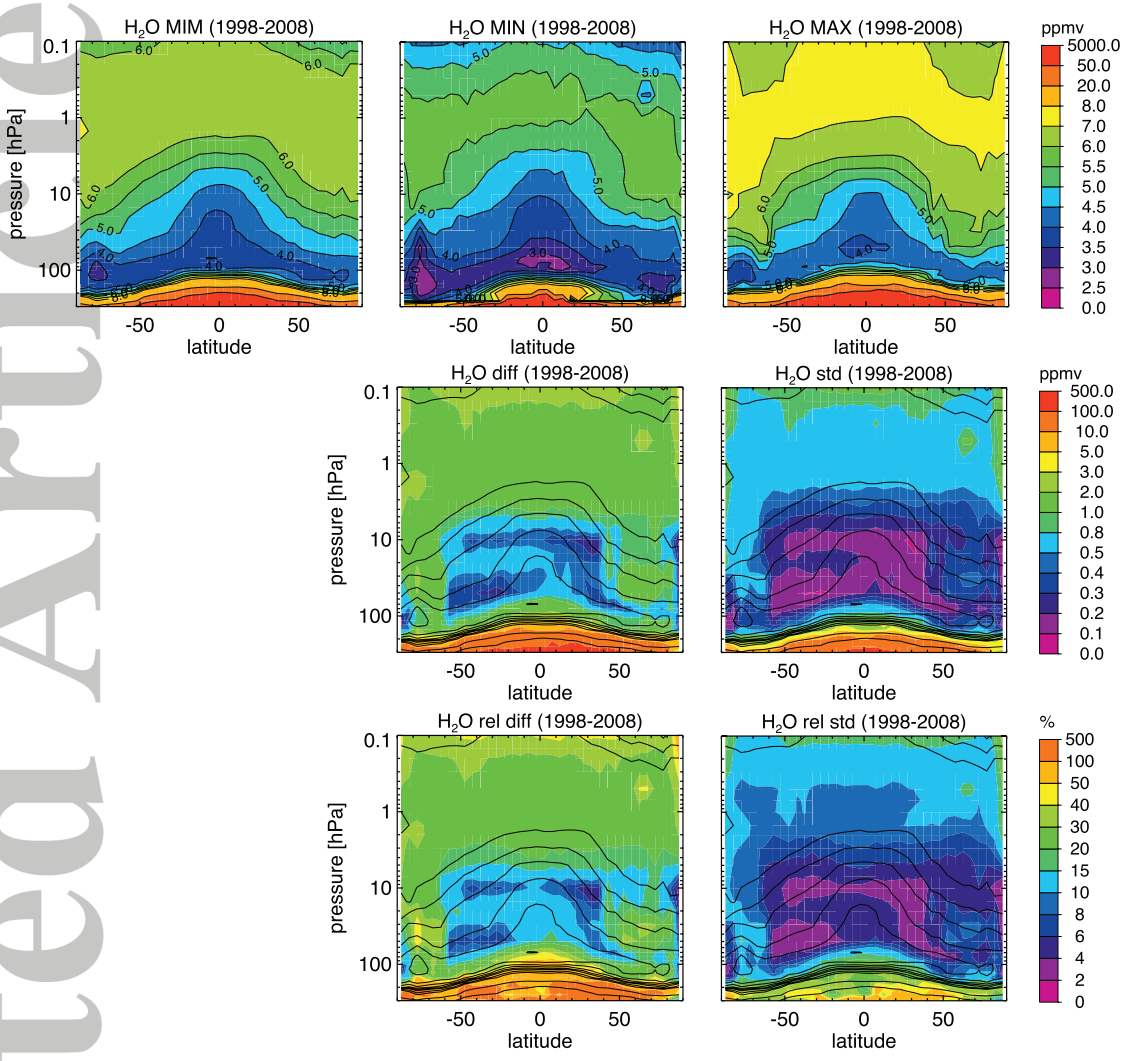

Figure 15. Summary of the water vapor annual zonal mean state for 1998-2008. Shown are the annual zonal mean cross-sections of the *MIM*, minimum (MIN) and maximum (MAX) water vapor values (upper row), the absolute differences (MAX-MIN) and absolute standard deviations (middle row), and relative differences and relative standard deviations with respect to the *MIM* (lower row). Black contour lines in the lower panels repeat the *MIM* distribution. Instruments included in the *MIM* are SAGE II, SAGE III, HALOE, POAM III, ACE-FTS, Aura-MLS, MIPAS(2), SAGE III, SMR(1), and SCIAMACHY. Note that SMR(2) and MIPAS(1) are not included in order not to bias the results towards one of the instruments.

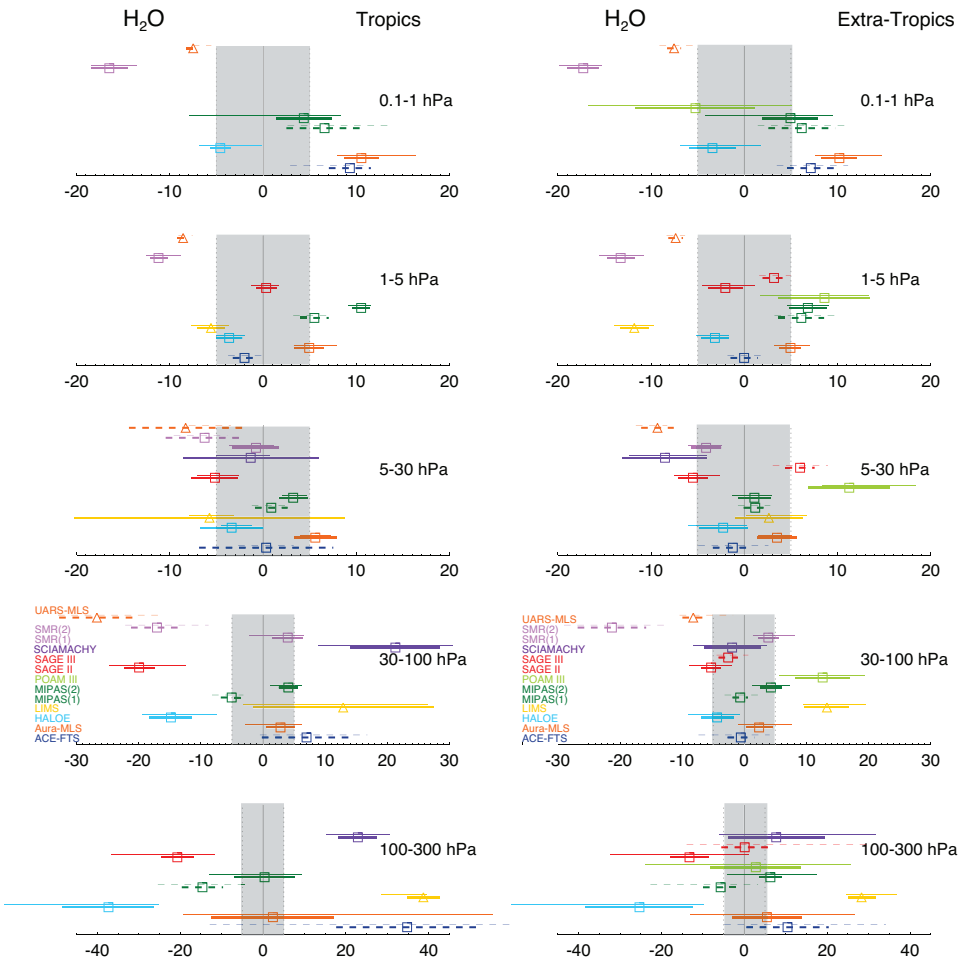

Figure 16. Inter-instrument differences in water vapor calculated for the tropics (20°S-20°N) (left) and extra-tropics (40°S-80°S and 40°N-80°N) (right) and for 5 different altitude regions from the UT up to the LM. Shown are the median (squares), median absolute deviations (MAD, thick lines), and the mean $\pm 1\sigma$ ranges (thin lines) of the relative differences between each individual instrument and the *MIM* calculated over a given latitude and altitude region. The reference period is 1998-2008 and based on the results of Section 4.1.3. Triangles indicate medians of instruments that are obtained outside of the reference period: LIMS and UARS-MLS are shown with respect to the instrument means of SAGE II and HALOE based on comparisons for 1978-1990 and 1991-1993, respectively.